Network
2020

Unlocking Commercial Opportunities

From 4G Evolution to 5G

GSMA

The GSMA represents the interests of mobile operators worldwide, uniting nearly 800 operators with more than 250 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and internet companies, as well as organisations in adjacent industry sectors. The GSMA also produces industry-leading events such as Mobile World Congress, Mobile World Congress Shanghai and the Mobile 360 Series conferences.

For more information, please visit the GSMA corporate website at **www.gsma.com**. Follow the GSMA on Twitter: **@GSMA.**

Network 2020

The GSMA's Network 2020 programme works with the mobile industry to deliver a collaborative and coordinated way of speeding up the development of all-IP communications.

The programme aims to produce a world in which all-IP global networks let mobile users enjoy seamless, secure and enriched communications on any device.

For more information, please visit the Network 2020 website at **www.gsma.com/network2020**

Contents

Executive Summary 2

1 Introduction 5
1.1 Scope 6
1.2 Definition of Terms 7
1.3 Document Cross-References 8

2 The Evolution to 5G 9
2.1 An Industry in Evolution 10
2.2 5G Business Environment 11

3 The 4G Evolution Business Opportunity 12
3.1 4G Evolution Opportunity Framework 13
3.2 Service Deployment Categories 14
3.3 Enhanced Mobile Broadband 14
3.3.1 Growing Data Demand 14
3.3.2 Traditional Mobile Broadband 15
3.3.3 Telecommunication Services 15
3.4 Massive Internet of Things and Critical Internet of Things 16
3.4.1 Business Model Implications 16
3.4.2 Connectivity Revenues 17
3.4.3 Beyond-Connectivity Revenues 17
3.5 Optimised Services 17
3.5.1 Pushing the Boundaries of Service Innovation 17
3.5.2 Broadcast Services 18
3.5.3 Vehicle to Everything Services 19
3.6 Developments Beyond 4G Evolution 20

4 4G Evolution Technical Strategic Roadmap 21
4.1 Technology for Accessing the New Business Opportunities 22
4.1.1 The 4G Evolutionary Path 22
4.1.2 4G Technical Enablers 23
4.2 Network Function Virtualization and Software Defined Networks 23
4.2.1 Technology Overview 23
4.2.2 NFV and SDN Deployment and Benefits 23
4.3 Mobile Edge Computing 29
4.3.1 Technology Overview 29
4.3.2 MEC Deployment and Benefits 29
4.4 Internet of Things 32
4.4.1 Technology Overview 32
4.4.2 IoT Deployment and Benefits 32
4.5 Evolved Broadcast Capabilities 33
4.5.1 Technology Overview 33
4.5.2 Broadcast Technologies Deployment and Benefits 34
4.6 Proximity Services 36
4.6.1 Technology Overview 36
4.6.2 Proximity Services Deployment and Benefits 36
4.7 Traffic Management and Quality of Service 37
4.7.1 Technology Overview 37
4.7.2 Traffic Management Deployment and Benefits 37
4.8 IP Multimedia Subsystem 39
4.8.1 Technology Overview 39
4.8.2 IMS Deployment and Benefits 39
4.9 Further Carrier Aggregation 41
4.9.1 Technology Overview 41
4.9.2 Carrier Aggregation Benefits and Requirements 41
4.10 Higher Order MIMO 42
4.10.1 Technology Overview 42
4.10.2 MIMO Benefits and Requirements 42
4.11 Heterogeneous Networks 43
4.11.1 Technology Overview 43
4.11.2 Heterogeneous Networks Benefits and Requirements 45

5 Conclusion and Recommendations 51

Annex A Document Management 53
A.1 Document History 53
A.2 Other Information 53

Executive Summary

The telecommunications industry is preparing to embark on the transition to the fifth generation (5G) of mobile technology. 5G promises a leap in technological capabilities, unleashing new opportunities to digitise more segments of everyday life. For operators, 5G is an opportunity to make a big push beyond connectivity and capture value from at least the Internet of Things (IoT), optimised services and mobile broadband.

In the 5G era, operators would need to explore new market routes beyond selling simply access and connectivity, and they would also need to nurture a non-B2C business in addition to the promise of the average revenue per user (ARPU) growth from B2C services that has underpinned previous generational shifts.

Between now and 2020, the year when the 5G system is expected to be commercially available, the technology advancement of the 4G system will continue unabated making it possible to provide existing services more efficiently and start creating a market for emerging applications (e.g. automotive, real-time gaming, personal cloud, sensor networks, remote health monitoring).

Mobile operators can start exploiting commercial opportunities by evolving their 4G network investments between now and 2020 in three areas:

1. **Massive IoT and critical IoT, mobilising vertical industries securely**
   Conservative estimations suggest that by 2025, the number of IoT devices will be more than double the number of personal communication devices. As the ecosystem grows, it will become clear that IoT is about more than connectivity, as this will only realise a small portion of addressable mobile revenues.
2. **Optimised services, making real-time communications real**
   A particular class of services earmarked in the available 5G literature for receiving a boost in usage are broadcast services. Operators could offer distributed content delivery network broadcasting, e.g. software updates or in-app advertisements, delivering live coverage of sports or music events. Where feasible, operators can become active distributors of infotainment services and directly monetise this opportunity.
3. **Mobile broadband, the perception of infinite bandwidth**
   Mobile broadband is the core business of operators aimed at the mass market and operators' data revenues have soared in recent years. Customers and operators are already demanding faster and more reliable access in the pre-5G era.

As operator communications services will coexist with numerous other services from alternative providers, it is unlikely that the current business model reliant on direct monetisation will be suitable in the future. The expectation is that retail charging will move towards a bundling of basic services (advanced or high QoE (Quality of Experience) service still attracting a premium), and that new revenue will be generated from providing dedicated virtual networks enabled by network slicing to high-value customers (e.g. enterprise, vertical industries, video streaming service providers, etc.).

A number of key emerging technologies in the 4G evolution roadmap have been identified as enablers for unlocking earlier the 5G commercial opportunities at radio level, core network level and system level.

In response to changing market realities up to 2020, the operators' network will have to support a number of existing and novel use cases often characterised by quite different requirements and business models. To be able to address these challenges, network slicing, the ability to create a virtual instance of the network optimised for the delivery of each type of service, is expected to become a key design feature of the service architecture. This functionality is enabled by Network Function Virtualization and Software Defined Networks.

The IoT ecosystem is projected to connect billions of devices (e.g. in areas of smart metering, automotive, transportation, smart cities and agriculture). Standards advances in Narrowband IoT and Long Term Evolution (LTE) Machine Type Communications (3GPP defined technologies) will enable operators to support the deployment of massive IoT devices in the licensed spectrum and allow them to insert themselves in the IoT value chain.

The ever growing popularity of non-linear multimedia content can be addressed through the introduction of Mobile Edge Computing (MEC). By gaining the ability to move content closer to the user, to be more responsive to the changes in radio conditions as well as to exploit evolved Multimedia Broadcast Multicast Service (eMBMS), operators will be able to introduce novel broadcast services as well as a distributed content delivery network.

The increasing demand for throughput, capacity and improvements to indoor coverage before 2020 can be fulfilled by deploying advanced technologies such as Carrier Aggregation, higher order Multiple Input Multiple Output (MIMO) and heterogeneous networks.

As various stakeholders and partners start providing detailed technical elements of the 5G ecosystem, it will become evident that an evolved 4G network will play a major role. Therefore, while the short- to mid-term evolution roadmap identified by the GSMA intends to allow to capture additional value earlier, the advent of the 5G system is not expected to devalue the investments made in 4G; on the contrary, these will make the mobile operator network 5G-ready.

1

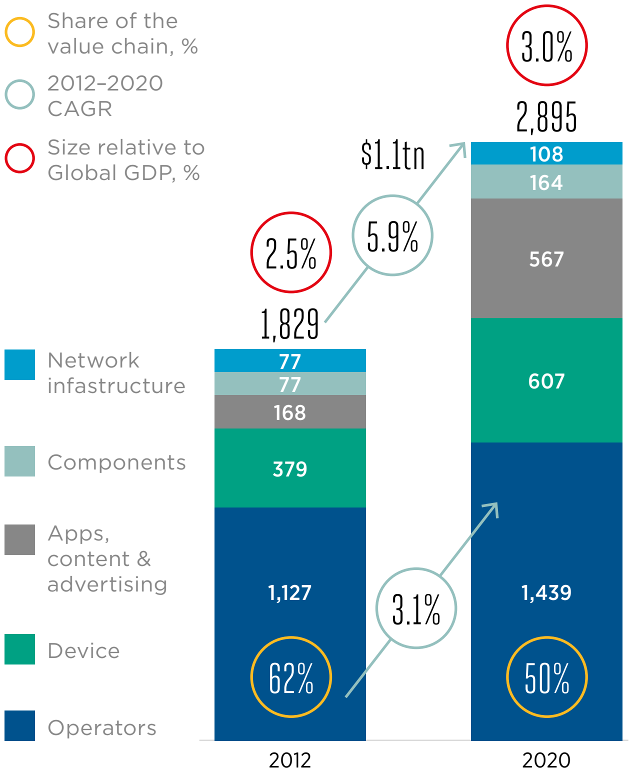
Introduction

Mobile telecommunications has had both a phenomenal and transformational impact on society. Starting from the earliest days of 1G analogue phones, every subsequent generational leap has brought huge benefits to customers and industry stakeholders and propelled ongoing digitisation of more and more segments of the global economy. 5G, the next generational leap, will continue this trend, building on the achievements of 4G and creating more opportunities for customers and industry stakeholders by unleashing higher capabilities that will spur innovation across society. Thanks to these advances, the GSMA predicts that the value of the mobile ecosystem, relative to global GDP, will rise from 2.5% in 2012 to 3% in 2020 when 5G adoption should be on the cusp of taking off.

For operators, the quest for faster and higher-capacity mobile broadband, plus the opportunities from IoT, will remain the key drivers for 5G. Operators would also need to expand the non-B2C business in 5G in addition to the promise of ARPU growth from B2C services that has underpinned previous generational shifts.

Source: GSMAi, Machina, Strategy Analytics, IHS Electronics & Media, Ovum, Gartner, Yankee, Value Partners analysis

Figure 1: The growing value of the non-M2M mobile ecosystem



Several industry groups, research labs, government institutions and academic centres are working to define the technology for 5G [1-8]. Much of this activity is driven by the need for faster and higher-capacity networks to support the boom in data usage and the anticipated explosion in the number of connected devices. However, the business considerations for 5G are also important as they underpin the decisions for spectrum auctions and network investments required to make 5G a reality.

The underlying commercial reality is that a number of services provided by operators at the time when 5G has reached a significant market share can be achieved with ongoing network evolution in 4G. This therefore provides guidance on how to capture both the evolutionary and revolutionary opportunities from 5G: basic services will be provided over an evolved 4G network while more advanced services will be carried by the new 5G system. This service provision model will be complemented by reconsideration, where necessary, of the business model/structure of the mobile industry.

1.1 Scope

This paper provides guidance to mobile operators on the commercial and technical opportunities towards 2020. As such, it frames the discussion by blending the technology developments together with the commercialisation opportunities from developments in the mobile industry. It takes into account the existing literature, and provides, as far as possible, a global view on technical, regulatory and commercial aspects that will characterise the mobile networks in 2020. The goal is to help operators to shape their strategy considering both new business opportunities unlocked by the advanced capabilities and how, through evolution of the 4G network prior to the deployment of 5G, operators can start benefiting from these opportunities.

Section 2 situates the expected business environment within the context of the evolutionary and generational changes in the mobile industry.

Section 3 explores the future business environment, focusing on the three major opportunities from Massive IoT, Optimised services and Enhanced Mobile Broadband.

Section 4 provides an overview of the technical enablers that can enrich the 4G system and outlines a short-/mid-term technical strategic roadmap towards a 5G system.

Conclusions and recommendations are given in Section 5.

1.2 Definition of Terms

Term	Description
3GPP	3rd Generation Partnership Project
AAA	Authentication Authorisation and Accounting
AMBR	Aggregate Maximum Bit Rate
ANDSF	Access Network Discovery and Selection Function
AP	Access Point
API	Application Programming Interface
APN	Access Point Name
ARP	Allocation and Retention Priority
ARPU	Average Revenue Per User
ATBC	Aggregated Transmission Bandwidth Configuration
BM-SC	Broadcast Multicast Service Centre
CA	Carrier Aggregation
CAGR	Compound Annual Growth Rate
CAPEX	Capital expenditure
CC	Component Carriers
CDN	Content Delivery Network
CDR	Charging Data Record
CEI	Customer Experience Intelligence
CN	Core Network
COTS	Commercial Off The Shelf
CPU	Central Processing Unit
CRE	Cell Range Expansion
CSFB	Circuit Switched Fallback
D2D	Device to Device
DC	Data Centre
DC	Dual Connectivity
DL	Downlink
DPI	Deep Packet Inspection
DRX	Discontinuous Reception
DVR	Digital Video Recorder
EC-GSM	Extended-Coverage GSM
eMBMS	evolved MBMS
eNB	eNodeB
EPC	Evolved Packet Core
ePDG	evolved Packet Data Gateway
EPS	Evolved Packet System
FDD	Frequency-Division Duplexing
feICIC	Further Enhanced Inter-cell Interference Coordination
FMO	Future Mode Operations
GBR	Guaranteed Bit Rate
GDP	Gross Domestic Product
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
GTP	GPRS Tunnelling Protocol
HA	High Availability
HeNB	Home eNodeB

Term	Description
HSPA	High Speed Packet Access
IaaS	Infrastructure as a Service
ICS	IMS Centralized Services
IMS	IP Multimedia Subsystem
IMT	International Mobile Telecommunications
I/O	Input/Output
IoT	Internet of Things
IPR	Intellectual Property Rights
IPX	Internetwork Packet Exchange
ITU	International Telecommunications Union
KPI	Key Performance Indicator
KQI	Key Quality Indicator
LAA-LTE	Licensed Assisted Access LTE
LTE	Long Term Evolution
LTE-A	LTE Advanced
LTE-WLAN	LTE Wireless Local Area Network
M2M	Machine to Machine
MANO	Management and Orchestration
MBMS	Multimedia Broadcast / Multicast Service
MBR	Maximum Bit Rate
MBSFN	Multicast-Broadcast Single-Frequency Network
MCE	Multi-cell/multicast Coordination Entity
MEC	Mobile Edge Computing
MeNB	Master eNB
MIMO	Multiple Input Multiple Output
MSC	Mobile Switching Centre
MT	Mobile Termination
MTC	Machine Type Communications
MU-MIMO	Multi-User MIMO
MVNO	Mobile Virtual Network Operator
NB-IoT	Narrow Band Internet of Things
NFV	Network Function Virtualization
NGMN	Next Generation Mobile Networks
NNI	Network-Network Interface
NSWO	Non-seamless Wi-Fi offload
OEM	Original Equipment Manufacturer
OPEX	Operating expenditure
OSS	Operations Support Systems
PC	Personal Computer
PCRF	Policy and Charging Rules Function
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PLMN	Public Land Mobile Network
PMO	Present Mode Operations
PMR	Professional Mobile Radio
PRB	Physical Resource Blocks

Term	Description
ProSe	Proximity Services
PSM	Power Saving Mode
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RAB	Radio Access Bearer
RAN	Radio Access Network
RAT	Radio Access Technology
RF	Radio Frequency
RNIS	Radio Network Information Services
RRC	Radio Resource Control
RRH	Radio Remote Head
RRM	Radio Resource Management
SaMOG	S2a Mobility over GTP Gateway
SC-PTM	Single Cell Point To Multipoint
SDL	Supplemental Downlink
SDN	Software Defined Network
SDO	Standards Developing Organisation
SeNB	Secondary eNB
SEP	Standard Essential Patent
SGW	Serving Gateway
SIP	Session Initiation Protocol
SIR	Signal-to-Interference Ratio
SLA	Service Level Agreement
SU-MIMO	Single-User MIMO
TCO	Total Cost of Ownership
TDD	Time-Division Duplex
TOF	Traffic Offload Function
TTM	Time To Market
UE	User Equipment
UHD/4K	Ultra High Definition
UL	Uplink
UMTS	Universal Mobile Terrestrial System
V2V	Vehicle-to-Vehicle
V2I	Vehicle to Infrastructure Unit
V2P	Vehicle to Pedestrian
V2X	Vehicle-to-Everything
ViLTE	Video in LTE
VNF	Virtualized Network Function
VoLTE	Voice over LTE
WAN	Wide Area Network
WebRTC	Web Real Time Communications
Wi-Fi	Wireless Fidelity
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless LAN
WRC	World Radio Conference

1.3 Document Cross-References

Ref	Document Number	Title
[1]		“5G White Paper 1.0”, NGMN
[2]		“5G New Wave Towards Future Societies in the 2020s”; 5G Forum
[3]		“LTE and 5G Innovation: Igniting Mobile Broadband”; 4G Americas
[4]	3GPP TR 22.891	“Feasibility Study on New Services and Markets Technology Enablers”; 3GPP
[5]		“5G Vision, Requirements and Enabling Technologies”; 5G Forum
[6]		“5G Vision and Requirements”; IMT-2020 (5G) Promotion Group
[7]		“Understanding 5G: perspectives on future technological advancements in mobile”; GSMA
[8]		“5G Vision”; European Commission’s 5G PPP
[9]	3GPP TS 23.246	“Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description”
[10]	3GPP TS 25.346	“Introduction of the Multimedia Broadcast/Multicast Service (MBMS) in the Radio Access Network (RAN); Stage 2”
[11]	3GPP TS 29.165	“Inter-IMS Network to Network Interface (NNI)”
[12]	GSMA PRD IR.95	“SIP-SDP Inter-IMS NNI Profile”
[13]	3GPP TS 36.300	“Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2”
[14]	3GPP TS 36.101	“Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception”
[15]	3GPP TS 22.278	“Service requirements for the Evolved Packet System (EPS)”
[16]	GSMA CLP.03	“IoT Device Connection Efficiency Guidelines”
[17]	3GPP TR 45.820	“Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT)”
[18]	3GPP TR 36.888	”Study on provision of low-cost Machine Type Communications (MTC) User Equipment’s (UEs) based on LTE”
[19]	3GPP TS 23.203	“Policy and charging control architecture”
[20]		“The Economic Benefit of Broadcast Offload in Mobile Data Delivery”, Rise Conseil & TDF
[21]		“Mobile-Edge Computing – Introductory Technical White Paper”, ETSI ISG MEC

2

The Evolution to 5G

2.1 An Industry in Evolution

The evolution between subsequent mobile technologies has responded to shortcomings identified in the previous generation. GSM (2G) addressed the poor spectral efficiency and security weakness of the analogue system (1G) and introduced roaming and packet wireless communications via the General Packet Radio Service (GPRS). The Universal Mobile Terrestrial System (UMTS) standards (3G) enhanced the data rate performances of GPRS, however they did not live up to the expectations of a true mobile internet experience.

It was only with the introduction of the High Speed Packet Access (HSPA) technology that consumers got a first taste of the world of opportunities provided by mobile broadband. Combined with more powerful devices, user friendly interfaces and the rise of app stores, it was inevitable that the design of 4G was heavily geared towards enhancing the mobile broadband experience.

This resulted in a packet switched only system delivering the “always on” promise, high throughput, low latency and realising the All-IP paradigm. This was possible by a combination of a new radio interface (LTE), a leaner architecture and a multi-access, packet switched only core network. Not surprisingly, across the world operators are typically reporting that 4G customers consume around double the monthly amount of data of non-4G users, and in some cases three times as much.

With a raft of new data intensive services gaining traction and consumers becoming more dependent on high speed internet, it is natural to conclude that the demand for mobile internet will continue to increase and that it will be one, if not the primary, design goal for 5G. The next generation mobile technology will need to provide higher throughput, lower latency and higher spectrum efficiency. The industry is already taking steps towards achieving these goals by evolving the existing 4G network with LTE-Advanced technologies such as Carrier Aggregation and MIMO plus the deployment of heterogeneous networks.

However, despite these enhancements to 4G, and as Table 1 shows, 5G has become necessary due to the natural progression to faster and higher-capacity broadband internet. Plus the need to capture value from the massive IoT opportunity, address the limited flexibility to support bespoke services across industry verticals, and develop next generation services that are not achievable with 4G networks.

Table 1: Evolution of technology generations in terms of services and performance

Generation	Primary services	Key differentiator	Weakness (addressed by subsequent generation)
1G	Analogue phone calls	Mobility	Poor spectral efficiency; major security issues
2G	Digital phone calls and messaging	Security, roaming, mass adoption	Limited data rates – difficult to support internet/e-mail demand
3G	Phone calls, messaging, data	Better internet experience	Real performance failed to match hype; failure of WAP for internet access
3.5G	Phone calls, messaging, broadband data	Broadband internet, applications	Tied to legacy, mobile-specific architecture and protocols
4G	All-IP services (including voice, messaging)	Fast broadband internet, lower latency	Not optimised for IoT scaling; limited flexibility to support bespoke services across industry verticals; inadequate for next generation services
5G	All-IP services, new technology sectors, verticals and end-users	Faster and higher-capacity broadband internet, lower (real time) latency, multi-access, multi-layered	

2.2 5G Business Environment

While 5G will ultimately be a technological leap, for industry stakeholders, the road to 5G will be partly shaped by ongoing trends and market realities. For example, the traditional role of mobile operators as both network access and services providers has evolved in an all IP environment with many other players also providing services to end users. Likewise, the primary role of operators as network access providers is no longer sacrosanct in an environment of widespread proliferation and usage of Wi-Fi, other small cell networks and potentially other macro cell networks such as those enabled by satellite, drones and balloons.

Other major trends that can be identified in the run-up to 5G deployment are:

- Usage will be dominated by data with revenues from voice and messaging expected to decline sharply.
- Users will have immersive communications at their fingertips regardless of access technology.
- There will be several tens of billions of connected devices. Connection identity mechanisms will evolve accordingly.
- New players enter the ecosystem, opening up new value chains (e.g. media companies, automotive industry).
- Competitive, regulatory and technological pressures may result in changes to the interconnection and the roaming business.

3

The 4G Evolution Business Opportunity

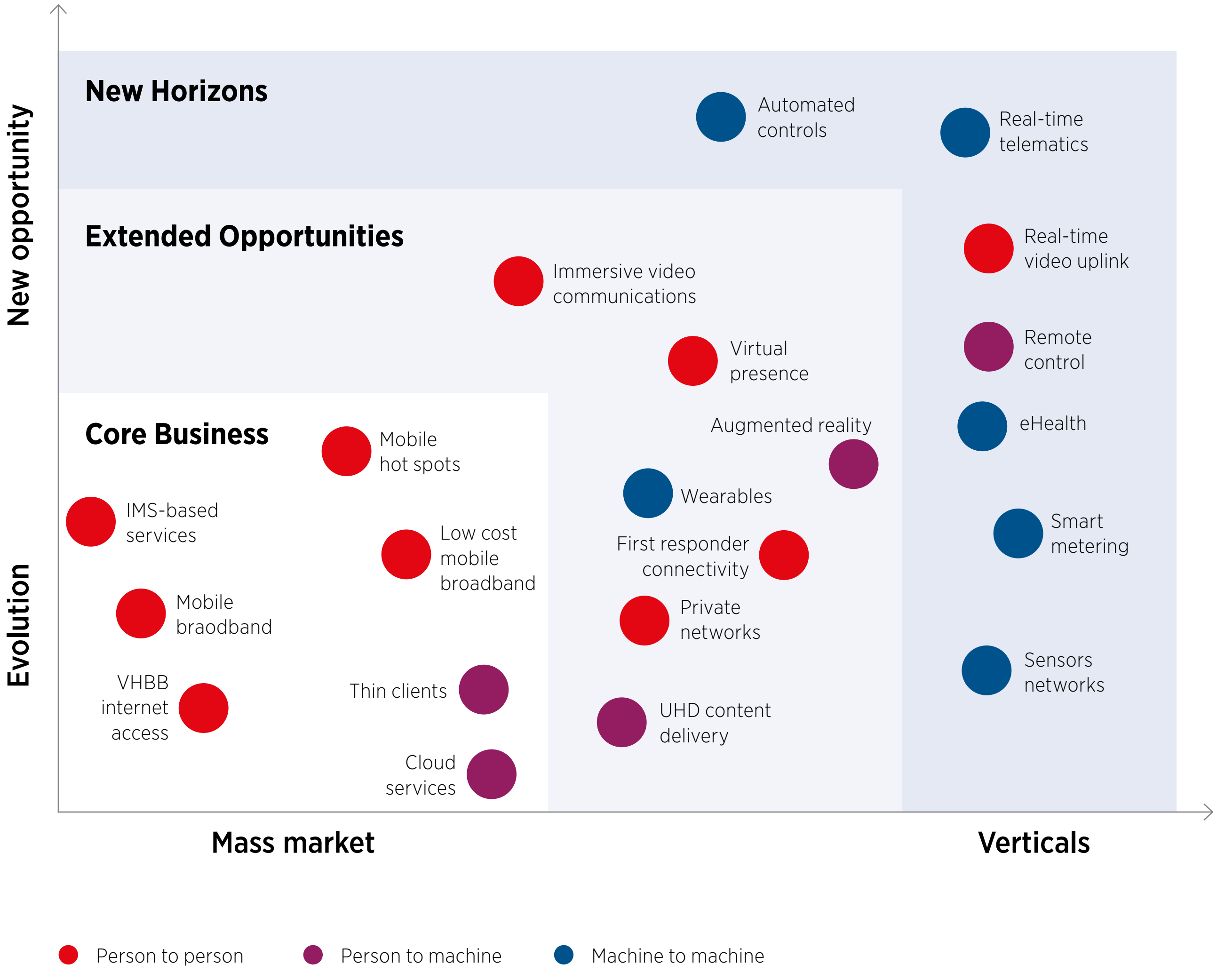
3.1 4G Evolution Opportunity Framework

There are two dimensions to future business opportunities as shown in Figure 2, below. Firstly, operators will evolve their current business while tapping into new opportunities that have been enabled by a more efficient technological framework. Secondly, while operators have historically focused on targeting the mass market, new capabilities will provide additional opportunities for operators to target niche segments within industry verticals. Some of these opportunities can be accessed by evolving the 4G network. However they will come to full fruition in a mature 5G system.

Whilst there has been some discussions on whether operators should offer only data connectivity in the future, it is pertinent to emphasize the need for a pragmatic business model that goes beyond selling only access and connectivity.

For example, IoT and broadcast opportunities are well suited to a model where operators leverage their distribution channels, billing infrastructure and customer relationships, to help businesses deliver services to their customers. Likewise, the focus on vertical industries with IoT and network slicing should boost the future share of operator revenues from the Business to Business segment.

Figure 2: The 5G opportunity framework



3.2 Service Deployment Categories

The GSMA expects that operators will have opportunities to expand their business in the following initial areas:

- Enhanced Mobile Broadband.
- Massive Internet of Things, Critical Internet of Things.
- Optimised Services.

3.3 Enhanced Mobile Broadband

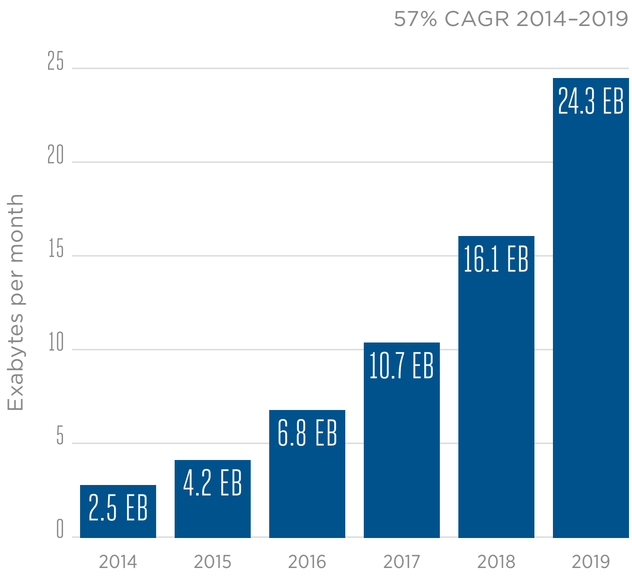
3.3.1 Growing Data Demand

Mobile broadband, in all its variants and together with its associated/complementary services, is the core business of operators aimed at the mass market. Accordingly, it will remain the main driving force behind the design and deployment of next generation mobile radio access technology as operators seek bigger and smarter data pipes to deliver services to customers.

This is not surprising as the advancements in devices capable of consuming higher bandwidth, a more and more connected society, cloud services, pervasive video and so on are driving a huge growth in data traffic. Cisco predicts that mobile IP traffic will reach 292 Exabyte's per annum by 2019, up from 30 Exabyte's in 2014.

Source: Cisco VNI Mobile, 2015

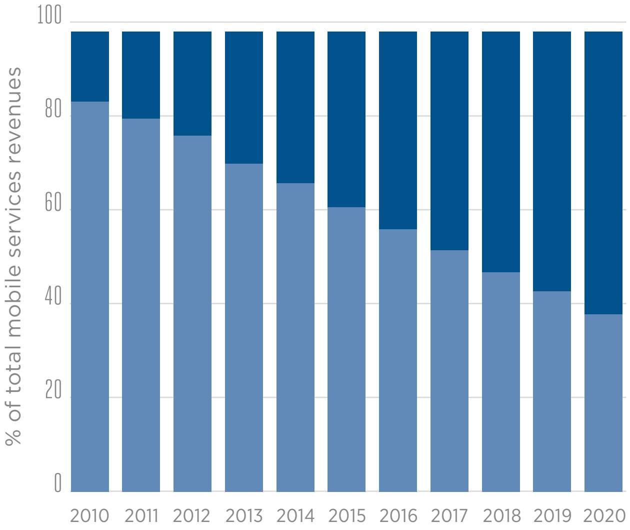
Figure 3: IP traffic growth prediction (1 Exabyte=10^{18} bytes)



Operators' data revenues have equally grown rapidly and the trend today is towards a rebalancing of tariffs towards mobile data and away from traditional voice and messaging. This rebalancing is nearly complete in post-paid-heavy advanced countries such as Japan and South Korea. But for operators in prepaid-heavy developing countries, the duration of this rebalancing is likely to be long and challenging.

Source: GSMA analysis

Figure 4: Tariff rebalancing – data revenues will dominate in the future



While tariff rebalancing from traditional telecommunication services towards mobile data provides the business case in the near term, there is still a debate on what the extent of a future commercial business case for mobile data is. The two key debates are:

1. Can operators sufficiently monetize mobile data growth so as to achieve a return on their network investment?
2. Is mobile data the only service that operators should offer in the future (the "bit pipe" scenario)?

In the context of future network evolutions, the GSMA evaluates these debates under two areas:

- Traditional mobile broadband.
- Telecommunication services.

3.3.2 Traditional Mobile Broadband

This is the often-talked about "bit pipe" scenario where operators are reduced to the role of bit carriers at the access level focussing on providing high bandwidth and low latency to consumers. It is the classic "utility" business model and is a predictable, profitable, but low-margin business.

Operators mostly rely on a usage-based business model for mobile broadband. Yet, with such a model, the availability of alternative access infrastructure (e.g. Wi-Fi and, potentially, satellites, drones, balloons) might constrain operators from adequately capturing value from data traffic growth as a future mobile network could become the secondary access option for customers, especially for large data downloads.

Furthermore, the opportunity for operators to use traffic management and differentiated Quality of Service (QoS) for specific services to derive additional value from mobile data may be constrained by Net Neutrality expectations[1].

3.3.2.1 Business Model Implications

The operator retail model for mobile broadband in the future is unlikely to be different from the status quo.

But in addition, an operator offering traditional mobile broadband may generate new revenue opportunities by providing dedicated virtual networks enabled by network slicing to target high value customers (e.g. enterprise, vertical industries, video streaming service providers such as Netflix and Amazon Prime, tiered consumer subscription level, etc.).

3.3.3 Telecommunication Services

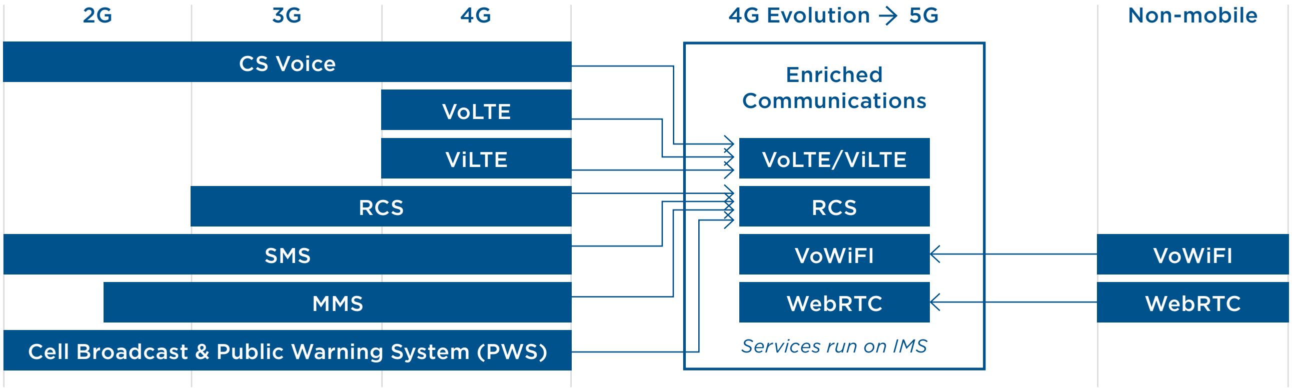
The traditional role of operators in providing communications services will continue in the future, even if operators are to share that role with other providers. Such service substitution, plus the complexity of providing emulation of circuit switched services using IP Multimedia Subsystem (IMS), has led to some uncertainty about the inclusion of interoperable operator voice and messaging services in post-4G standards.

However, the GSMA notes the following:

- **Regulatory obligation:** Most mobile licenses mandate the operator to provide communication services (often this is implicitly specified in requirements for Emergency Calling).
- **Manage the tariff rebalancing process:** Voice and messaging services will generate over $500bn in annual revenues in 2015 (GSMAi) and an industry decision not to offer them in the future will have an immediate and disastrous impact on many operators' tariff rebalancing strategy.
- **Asset utilisation:** There is huge ongoing investment in IMS and Voice over IP (VoIP) technologies, which will not be written off soon ($4.2bn in 2014; $4.9bn by 2019 according to IHS Infonetics).
- **Low marginal cost:** While it is acknowledged that the roll-out of Voice over LTE (VoLTE) was slower than expected and that part of the blame has to be apportioned to IMS, the expectation is that most mobile operators will have a fully functional VoLTE service by 2020.

From these observations, the GSMA proposes that 4G Evolution networks should continue to offer, and build on the full set of IMS-enabled communication services that are already available in 4G. The expectation is that these services, packaged as an enriched communications proposition to customers, will replace legacy communications services and will become the base assumption for future 5G networks.

Figure 5: Communications services evolution



---

[1] Many markets have not enacted Net Neutrality regulation and there is freedom to offer specialised services (i.e. access optimised for specific service characteristics) within the rules proposed in the United States of America and the European Union. Net Neutrality will remain a key public policy topic as consumers increasingly access IP-based services; but it does not preclude innovation in new services and commercial propositions.

3.3.3.1 Business Model Implications

Operator communications services in the future will coexist with numerous other services from alternative providers. Given this competitive landscape, it is unlikely that the business model, which is still reliant on direct monetisation (especially through rating or per-minute billing), will be suitable.

Therefore, while support of interoperable communications services could be considered a given, the expectation in terms of charging is that retail charging will move towards a commoditisation of basic services (advanced or high QoS services still attracting a premium).

3.4 Massive Internet of Things and Critical Internet of Things

The IoT opportunity continues to boom, driven in part by ongoing digitisation of more segments of the economy. Conservative estimates by analysts suggest that by 2025, the number of IoT devices will reach 35 billion, at least double the number of personal communication devices. These IoT devices will cover diverse use cases spanning from low-cost sensors transmitting sporadically small amounts of delay-tolerant data to in-vehicle devices requiring high bandwidth and stringent delay budgets through surveillance camera and wearable devices. Each of these Machine Type Communication devices will not only place different types of demands on the transport network, but will also have different addressing and security requirements. A special class of devices forming the critical IoT will require very low latency or high reliability or both.

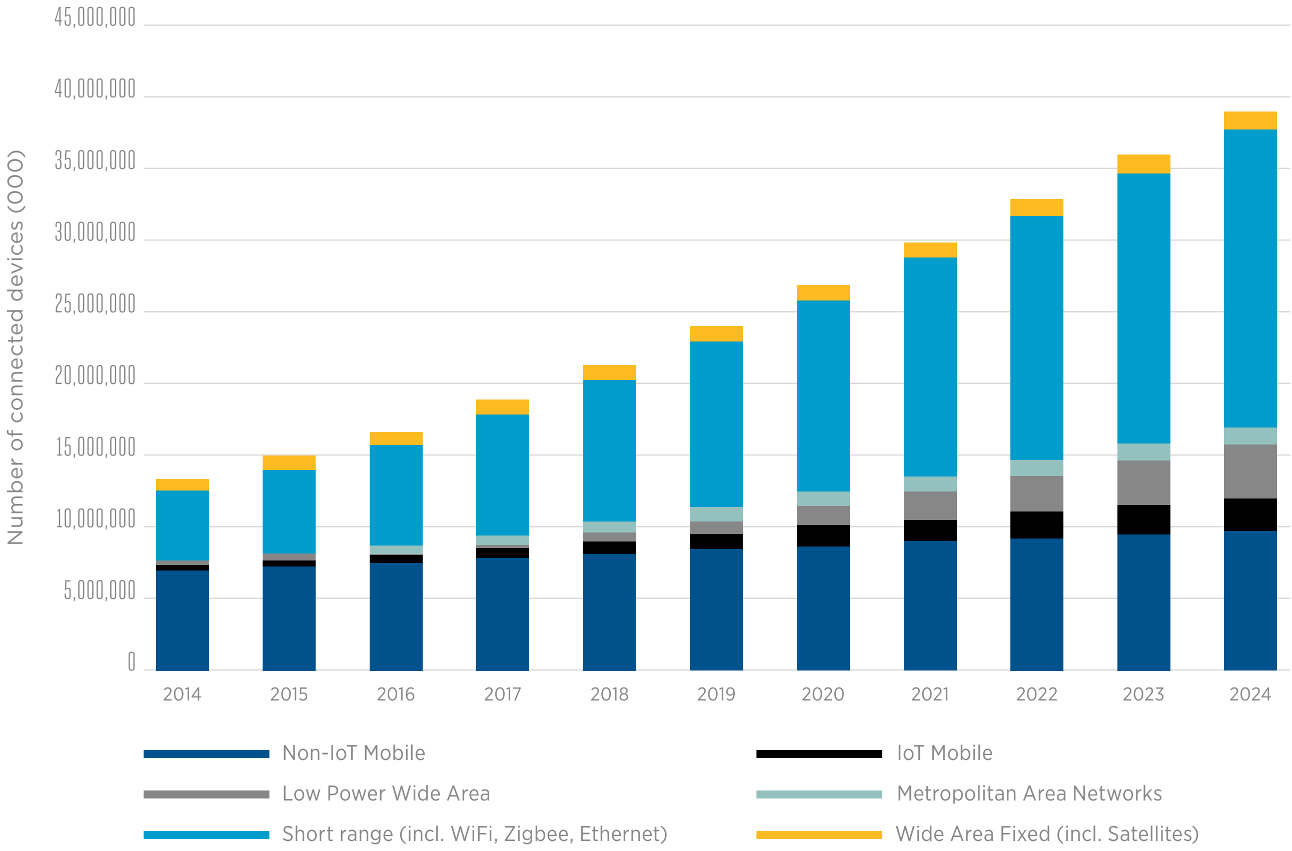
In order to cater for this wide variety of requirements that IoT services have, mobile operators will need to make full use of the capability to create virtual dedicated networks in order to deliver these services in an efficient and cost-effective manner.

3.4.1 Business Model Implications

New network capabilities will create more IoT opportunities for operators, in addition to the many IoT use cases that can and are being addressed using today's existing technologies. For operators, the IoT opportunity is primarily about many more connections, and about capturing value from parts of the IoT value chain beyond connectivity.

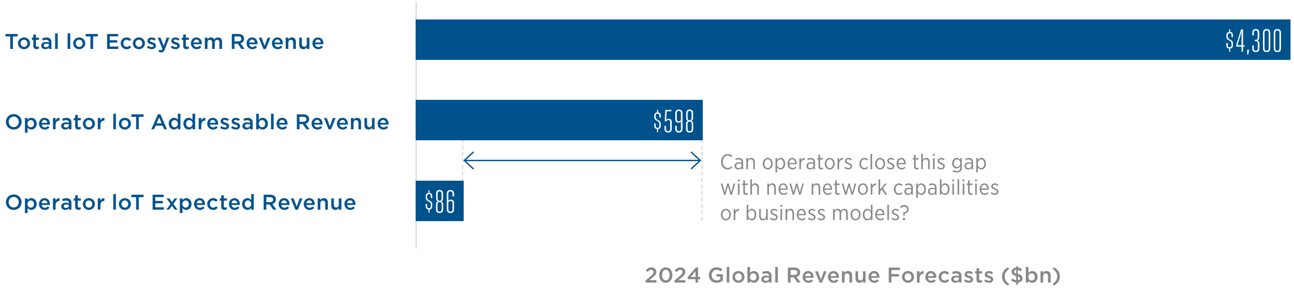
Source: Machina Research

Figure 6: Global connected devices growth



Source: Machina Research

Figure 7: The future IoT opportunity



3.4.2 Connectivity Revenues

Connectivity revenues represent the main revenue opportunity for operators in IoT and this opportunity will continue to grow as billions of additional connections are brought on stream. However, operators will not be the only providers of network access for the IoT ecosystem as the mobile network will coexist and compete with other access technologies such as Wi-Fi, Low Power Wide Area Networks and satellites

3.4.3 Beyond-Connectivity Revenues

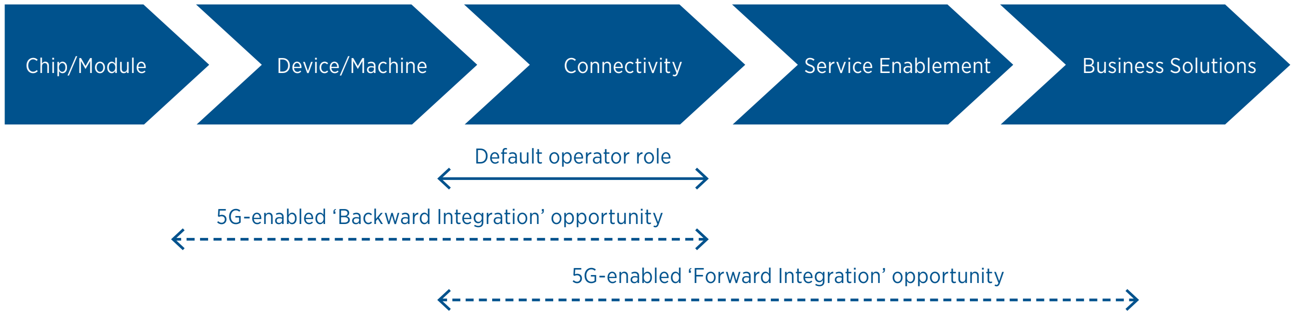
A typical example of the value chain for an IoT service is illustrated in Figure 9. 4G evolution should support operators' ongoing efforts to expand into segments other than being the provider of connectivity (e.g. taking advantage of the 5G system's capabilities to offer critical IoT services or productising/monetising security). A "Forward Integration" opportunity should see operators take on more roles in enabling IoT services, and where feasible, offering IoT business solutions. These could also include analytics/Big Data, real time control/telematics, and autonomous driving capabilities. Likewise, operators could backward integrate and take advantage of opportunities in new areas. Already some operators have begun expanding into this space with their home security solutions (e.g. AT&T's Digital Life).

3.5 Optimised Services

3.5.1 Pushing the Boundaries of Service Innovation

Continued improvements in speed and latency in 4G Evolution networks, and eventually 5G, will provide superior capabilities to nurture and support new services that cannot be otherwise supported.

Figure 8: Typical value chain for an IoT service



Some of the services being considered include:

- Tactile internet. Allowing the full range of human senses (not restricted to the auditory and the visual) to interact with machines and technology.
- Immersive communications (virtual/augmented reality) Technology that simulates physical reality, blurring the distinction between the physical world and the digital for the user.
- Telepresence. Enabling users to control machines remotely, participate in events that are taking place in remote locations, etc. via the use of virtual reality technologies.
- Industrial control systems. Controlling remotely located machinery used in industrial production based on input from devices located in the field.
- Remote control of vehicles and robots. Controlling vehicles and robots via the use of tactile internet and telepresense technologies.
- Connected vehicles. Vehicles equipped with internet access, cellular radio, radar and other communication links including an internal wireless local area network, enabling vehicles to communicate with each other and their surroundings.
- Holograms. A 3D image of an object, person, etc. projected to a surface, which creates a floating effect.

3.5.2 Broadcast Services

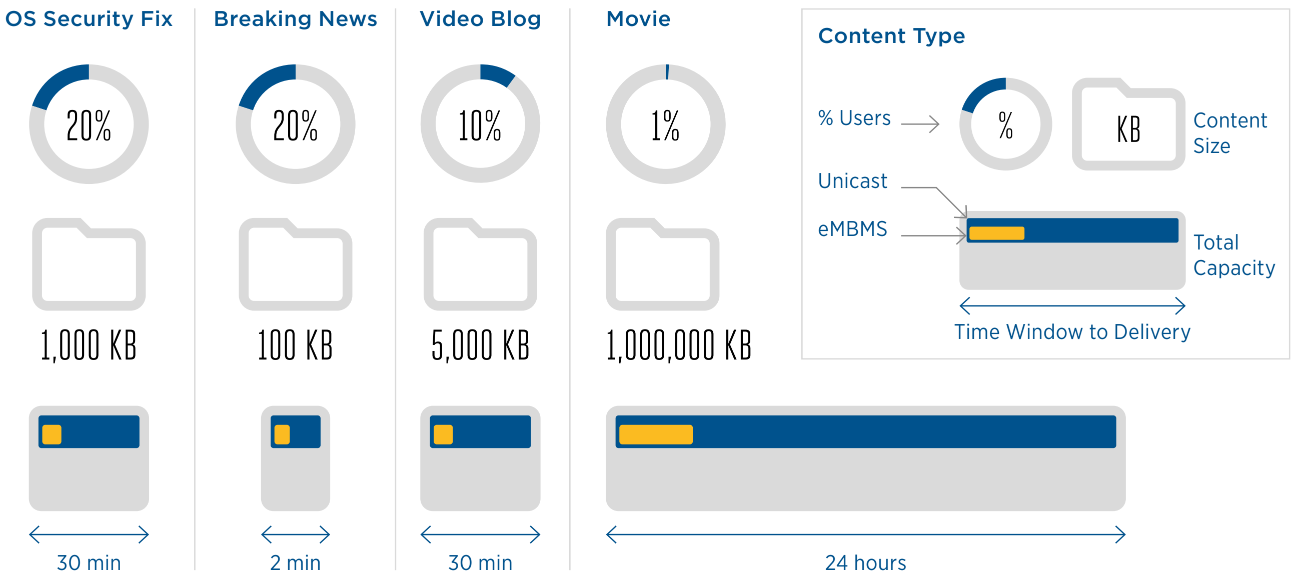
A particular class of services that is expected to receive a boost in usage is broadcast services. Compared to the point-to-point links of today's mobile broadband services, broadcasting or multicasting (where only authorised devices can successfully consume the broadcast contents) provide the most optimal means to distribute identical information to a large population of terminals geographically dispersed over a large area and across a number of different cells.

The evolved Multimedia Broadcast/Multicast Service (eMBMS) supports broadcast and multicast services in LTE networks over multiple cells and is one of the building blocks of a future 5G network. Additionally, Single Cell Point to Multipoint (SC-PTM) broadcast supports multicast services over single cell using LTE downlink shared channel allowing efficient radio utilisation and flexible deployment of number of applications. Mobile Edge Computing (MEC) may also be deployed to enable efficient use of network resources via the caching of broadcast/multicast content near the network edge prior to transmission.

The commercial future of broadcast in 4G evolution is primarily as an efficient Content Delivery Network (CDN) and, where feasible, opening new opportunities in media distribution. Three use cases are foreseen:

Source: Alcatel Lucent: https://techzine.alcatel-lucent.com/further-efficiencies-embms-preloading

Figure 9: Relative advantage of broadcast over unicast



1. **Optimally delivered services:** Amidst the frenzy over media streaming, it is worth noting that a mobile broadband network cannot provide high throughput for all subscribers in a coverage area at the same time, even with 4G evolution's superior spectral efficiencies. This is an important concern in markets with sparse fixed telecoms networks and Wi-Fi. Therefore, to maintain service quality and optimise the use of network resources, operators should use eMBMS to deliver selected content to their customers. Examples will include:
   - Live coverage of events (e.g. sports, festivals, concerts, etc.).
   - Updates for apps, operating systems, firmware and software.
2. **Infotainment services:** Where feasible and permissible by law, operators could become active distributors of broadcast infotainment services, such as TV and video. A corollary can be seen in the fixed telecoms industry where several operators (e.g. BT) have expanded their TV/video operations in order to support the business case for their fibre broadband services.
3. **Public information broadcast:** While radio and television remain the primary means of broadcasting public information, recent examples have shown the impact of coordinating public action through the internet and social media. Operator broadcast capability will be a necessary complement for public safety information (e.g. in disaster zones).

3.5.2.1 Business Model Implications

There is commercial opportunity for operators in broadcast services if the business model is right. This is particularly important to operators in developing countries where fixed telecoms infrastructure is sparse and where the cost of mobile data constrains the usage for many customers. The following commercial considerations should be made for broadcast services to become a commercial reality:

- **Device support for eMBMS or SC-PTM and media recording/storage:** To achieve scale quickly and ensure that any commercial strategy for broadcast services is not handicapped by lack of devices, it is necessary for future 4G devices to support eMBMS by default. Another requirement should be for such devices to reserve some memory (aspire for at least 10% of available memory) to be used for recording and storing media content. This should work as a form of Digital Video Recorder (DVR).
- **Efficiency benefits:** Rather than expecting customers to pay directly for the broadcast service, operators should capture value in the efficiency gains and improved customer experience that broadcast provides. Analysis by Rise Conseil & TDF [20] shows that broadcast offload could result in savings of €1bn in France in 2019, about 15% of the annual operator capital expenditure (CAPEX) in the market.
- **Revenue from carriage:** Operators may charge fees to businesses such as TV/video companies and software companies to distribute their content to end users via broadcast. This is a classic B2B2C opportunity.
- **New media revenues:** Where feasible, operators can set up their own media, operation which they can directly monetise.

3.5.3 Vehicle to Everything Services

The motivation of V2X (Vehicle-to-Everything) communication is to improve travel safety, minimize environmental impact, improve traffic management and maximize the benefits of transportation to both commercial users and the general public. V2X offers great commercial opportunities for both vehicle and communication industries. Lots of investments from governments, academia, manufactures and vendors have been attracted into this new emerging business field.

V2X enables vehicles to communicate with other vehicles, infrastructure units and even pedestrians, which are also known as V2V, V2I and V2P separately. By using V2X communications, vehicles and more specifically drivers are able to sense the status of the surrounding vehicles, instruction from infrastructure unit and even acquire information on approaching pedestrians. Different kinds of pre-defined cooperation and notification V2X messages containing moving speed, moving direction and even path history, are exchanged between V2X transmitters and receivers. By analyzing the messages received, vehicles give the corresponding notification or warning to the drivers. Then drivers are able to be aware of impending danger and traffic status in advance, which could effectively reduce fatalities and serious injuries that result from crashes, or reduce negative effects from traffic congestions, fuel consumption and even CO2 emission.

The V2X feature will be introduced into 3GPP in Release 14 and the corresponding technical specifications are expected to be made available in first half of 2017.

3.5.2.2 Business Model Implications

V2X is a clearly-defined use case for D2D (Device to Device) communications. Its deployment topology is more suitable for low latency communications, a capability that would be highly valuable for future transportation systems.

Commercially, today's operator business model, which is reliant on data traffic with the macro network, will be inadequate for V2X services because there is yet no clear mechanism for paying for D2D communications. Instead, commercial V2X services will require a business model that compensates the provider of macro-level connectivity and coordination. Such macro-level coordination might be needed to allocate resources, resolve conflict and provide an audit trail.

3.6 Developments Beyond 4G Evolution

In the run up to 2020, a number of events will impact the mobile ecosystem:

- The traditional subscriber identity module role will undergo a significant transformation. The industry, led by the GSMA, is already developing a new model.
- World Radiocommunication Conference 2015 (WRC-15) and World Radiocommunication Conference 2019 (WRC-19) will determine the amount and frequency of the spectrum available to operators.
- The standardisation of 5G, now in its infancy, will mature with clear technological trends emerging from the end of 2017 onwards.

While it is clear that mobile operators will continue to offer and manage identities and addressing, provide secure access to the network and offer interoperable services, it is too early to determine how the business model will be affected.

The GSMA is working with the relevant stakeholders to help drive technical standards ensuring that mobile operators' requirements are fulfilled by the 5G system, for example, in roaming and interconnection, interoperable communication services, security, authentication, authorisation and identities.

4

4G Evolution Technical Strategic Roadmap

4.1 Technology for Accessing the New Business Opportunities

4.1.1 The 4G Evolutionary Path

The transition between 3G and 4G was driven primarily by the need for higher speed, more capacity and lower latency but the business model and use cases remained very much the same. Recent developments in the industry have now opened the opportunity for operators to address new areas of the market (e.g. V2X) as well as consider again certain opportunities that technology advancements have made more attractive (e.g. broadcasting, critical communications).

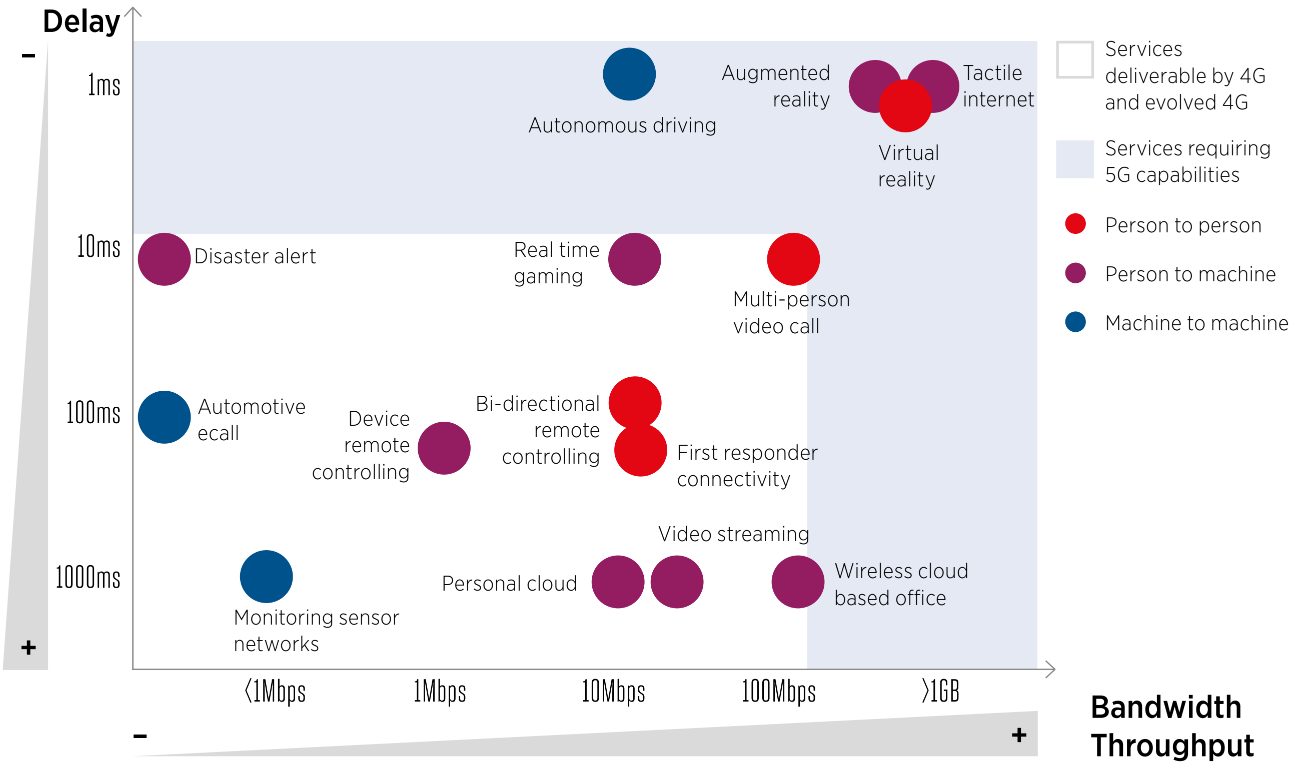
At the same time, it is commonly understood that there are classes of applications that rely on the superior 5G performances for latency (sub 1ms on the radio access), capacity (×100 the capacity of 4G) and peak data rate (20 Gbps / 10 Gbps for downlink and uplink respectively).

The GSMA's analysis paper issued in December 2014, "Understanding 5G: Perspectives on future technological advancements in mobile" [7] highlighted that a significant proportion of the new business opportunities can and should be accessible using an evolved 4G network (see Figure 11). This has crucial commercial implications since it means that 4G will continue to be used to support legacy services (including IMS-enabled communications services) and therefore maximises the return of investment in 4G, which the GSMA estimates to be in the tune of $1.7 trillion in the 10 years to 2020. Technically, it will be shown in the remainder of section 4 how the 4G system enhancements can place an operator in the position of delivering many of the services to be offered in the future.

An even more important commercial reality is the distribution of mobile connections in the near future. 5G will coexist with previous mobile generations, accounting, at global level, for about 10% of total non-machine-to-machine (non-M2M) connections by 2025, whereas 4G will account for 60% of connections[2]. Along the same lines, Juniper Research forecasts that 5G services will account for less than 10% of mobile services revenues by 2025. This reality has profound implications for the 5G business case: it validates the strategy to have 5G coexisting with 4G evolution as mobile operators will seek to capture new opportunities with 5G while maintaining service and business continuity for their existing operations.

Source: GSMA Intelligence

Figure 10: 5G supported services matrix



---

[2] In advanced countries the 5G penetration is expected to be significantly higher.

4.1.2 4G Technical Enablers

The remainder of section 4 analyses a number of technical enablers that are available to be deployed within the existing 4G network with the purpose of highlighting how they can help mobile operators to realise most of the business opportunities discussed in section 3.

Before 2020, both the 4G core network and the access network will undergo a transformation with new functionality being added and existing functionality enhanced.

4.1.2.1 Core Network Considerations

The Evolved Packet Core (EPC) was not designed to be access agnostic. Like 2G/3G packet core, EPC also has dependencies on the 3GPP RAN. For example, Evolved Packet System (EPS) bearers, QoS principles, Authentication Authorisation and Accounting (AAA) mechanisms in EPC are LTE specific. Similarly, the policy framework and mobility as defined in EPC are 3GPP specific and hence support 3GPP specific mechanisms.

EPC was developed by 3GPP to provide a tight coupling with legacy 2G/3G packet core. This approach allowed seamless service continuity while providing for coexistence with legacy systems. However, a tight coupling with legacy 2G/3G packet core came at a cost. In many instances, key legacy protocols (e.g. GPRS Tunnelling Protocol (GTP)) and concepts (e.g. Access Point Name (APN)) were carried forward to keep alignment with legacy systems. Even while carrying forward several of the legacy concepts and protocols, packet core complexity still increased in order to support interworking between 2G/3G and EPC packet core entities.

With time, other access technologies such as Wi-Fi had to be integrated with 3GPP access technologies such as LTE and because of the access specific nature of EPC interworking techniques were developed in EPC to work with non-3GPP access technologies to support mobility, QoS, AAA and policy. However, due to the complexity of interworking, these solutions have seen few commercial deployments.

Within the limits of this framework, however, EPC functionality is expected to be further evolved and this evolution is described in sections 4.2 through 4.8.

4.1.2.2 Radio Network Considerations

With regards to radio access, 3GPP is in the process of finalising the work started in Release 10 on LTE-Advanced, which will deliver the following:

- Increased peak data rates (downlink 3 Gbps, uplink 1.5 Gbps).
- Higher spectral efficiency (from a maximum of 16bps/Hz in 3GPP Release 8 to 30 bps/Hz in 3GPP Release 10).
- Increased number of simultaneously active subscribers.
- Improved performance at cell edges.

The enhancements / functionality enabling these goals are described in sections 4.9 through 4.11.

4.2 Network Function Virtualization and Software Defined Networks

4.2.1 Technology Overview

As discussed in section 3, unlocking new business opportunities will require the mobile operator network to be able to adapt to the communication needs of vertical services and this will pose the challenge of being able to adapt the traditionally monolithic design of the network to become multipurpose. Two technologies are often cited as key enablers for such practical and versatile architecture: Network Function Virtualization (NFV) and Software Defined Networks (SDN).

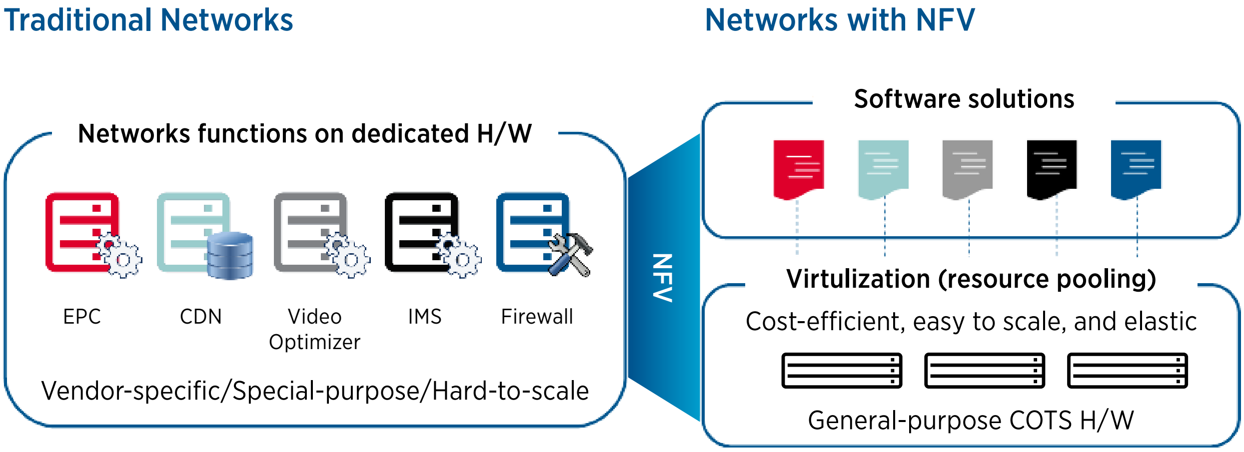
Besides the gains in terms of operating expenditure (OPEX) and elasticity (that is, the capability to dynamically allocate resources where they are most needed), NFV and SDN will empower mobile operators to optimise the usage of their physical resources by creating virtual end-to-end-networks specialised for the support of a service type.

4.2.2 NFV and SDN Deployment and Benefits

4.2.2.1 Architecture Description

Network Functions Virtualization decouples software of a given network function from its dedicated hardware and allows such software to run on commodity hardware. Figure 13 illustrates how NFV transforms the traditional networks into more cost-efficient and elastic networks.

Figure 11: Decoupling software from dedicated hardware using NFV

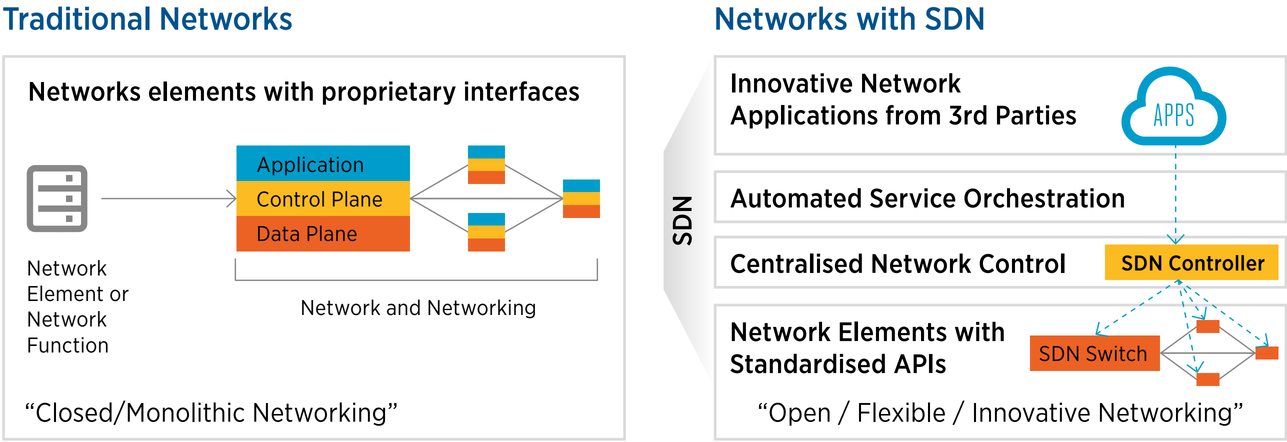


Software Defined Networking further decouples the control plane function of given software from its corresponding data (user) plane function. The control functions can then potentially be aggregated and provide northbound APIs for operators to dynamically and systematically control the network in real-time. Figure 14 illustrates how SDN transforms the traditional networks into more open, flexible, and innovative networks.

SDN together with NFV ultimately allows operators to control the network in a more dynamic, programmable and consistent manner.

The implementation of NFV and SDN present a number of short-term challenges to be noted. Network functions are currently tightly coupled and optimised to run on specialised hardware for better performance. When the network functions are decoupled in NFV to run on

Figure 12: Decoupling control plane from data plane for open, flexible, and innovative networking



commodity hardware, it is anticipated that there will be initially a performance degradation, as the software will no longer be able to take advantage of the specialised hardware. This shortcoming can be addressed with the use of various hardware accelerators, however, such use results in another form of hardware dependency. This observation emphasises and calls for a need to standardise an abstraction layer for the various hardware acceleration technologies in NFV.

Another key area which is still in its infancy in terms of a fully developed solution and standardisation, but crucial to realising the benefits of SDN and NFV, is orchestration. In essence, the role of the orchestrator is to deliver the real Operations Support Systems (OSS) value as it offers the following major functions:

- End-to-end provisioning.
- End-to-end resource management.
- Supporting service models driven by business metrics.

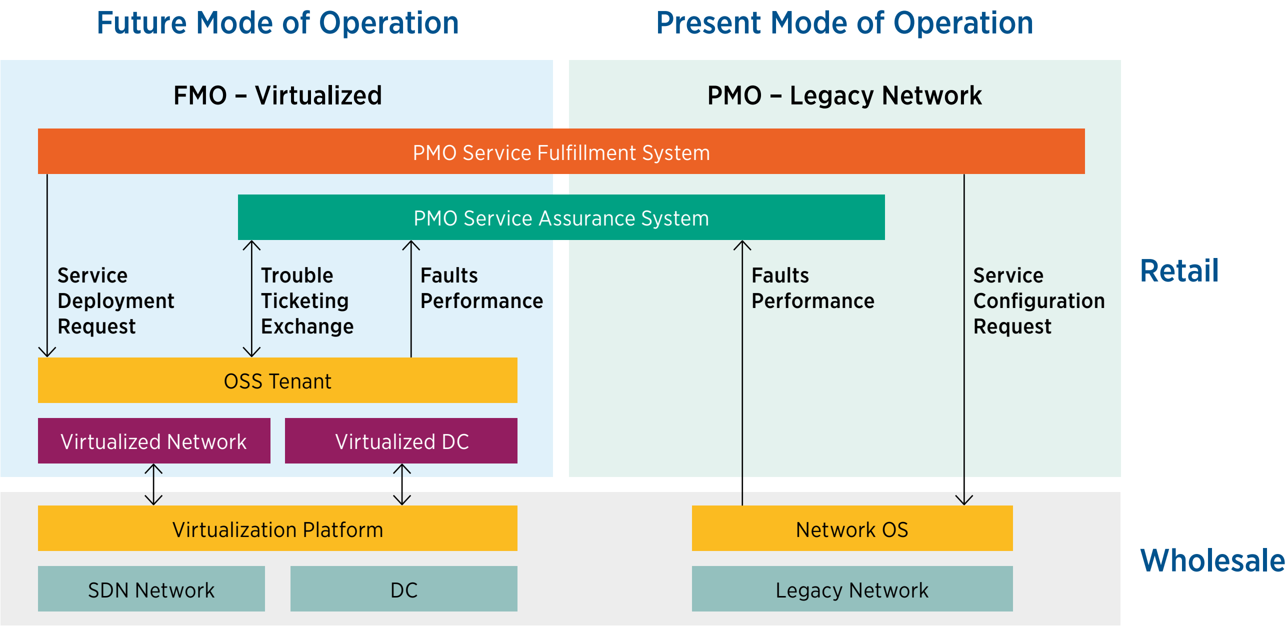
Service orchestration is at the heart of automating the future service lifecycle management (Future Mode Operations or FMO for short). The orchestration engine must be able to coordinate the activation of an SDN network across the transport layer and the data centre (DC) network layer to form a virtualized and distributed network with DC processing capabilities across all connected locations inclusive of various virtualization network functions.

Virtualized SDN/NFV networks management brings the following new challenges to an OSS system:

- Support of dynamic, policy-driven (near) real-time processes, root-cause and customer impact analysis.
- Tight connection between assurance and fulfilment processes.
- Intuitive and user-friendly service creation environment.
- Integrated analytics – based on real-time traffic measurement, customer behaviour, customer usage, etc.
- Service exposure via open APIs, creation of open business partner ecosystem.
- Support of automated operations for both physical and virtualized resources with limited (or without) human involvement.

With a network composed of physical resources, hop-by-hop troubleshooting has its challenges but is relatively easy. However, with virtualized resources, troubleshooting becomes even more challenging, because the physical-to-virtual mapping can change in real time. Future troubleshooting activity will need appropriate tools to ensure real-time mapping of virtualized functions and that associated physical resources are available. These tools should take into account service chaining and the impact of the full or partial movement of Virtualized Network Functions (VNFs) across data centres.

Figure 13: Service orchestration overview

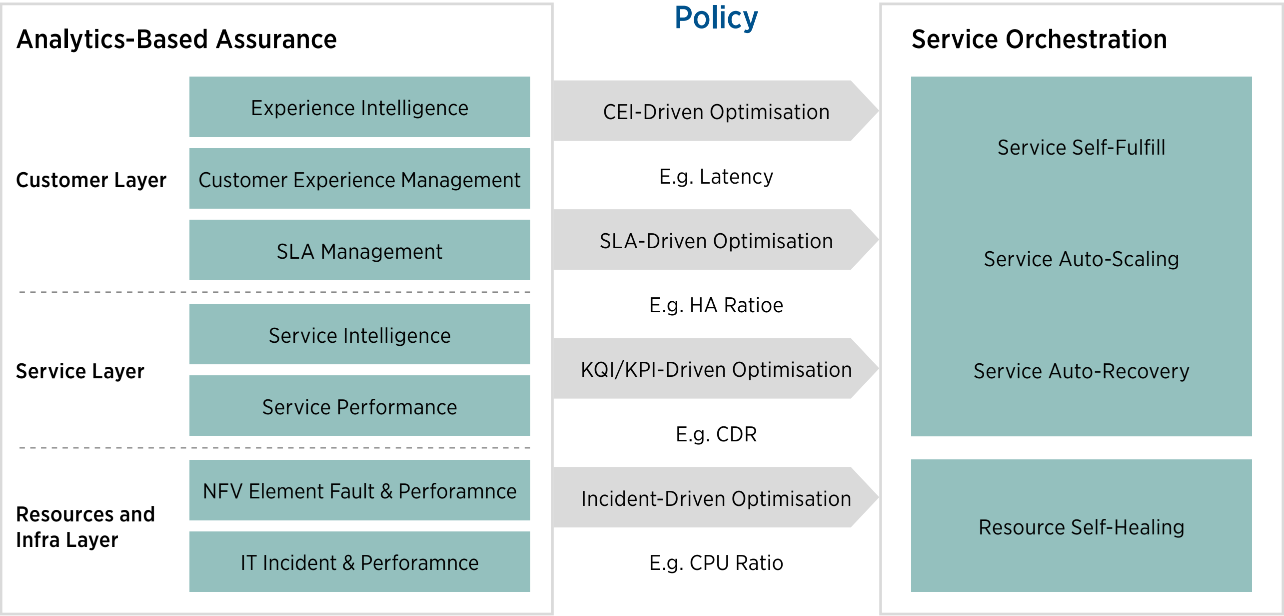


Another crucial element within the orchestration is the role of analytics capabilities. The ability to generate actionable events from collected data and to act automatically on those events to improve customer experience or Service Level Agreement (SLA) performance standards are essential to realise business value. By applying data extraction and data mining techniques to collected data, triggers are expected to be generated in real time. This information will subsequently need to go to the policy engine to send policy directions back into the orchestration layers to self-regulate, self-optimise, and relay important network status changes. This can be a challenging task using the tools available today.

Using SDN/NFV, data velocity, volume and variety increases, but more value is realised from better insight into Quality of Experience (QoE), customer management and end-to-end management.

A real-time analysis platform can work on four categories of triggers for optimising and taking corrective actions as shown in Figure 16. These categories are Event/Incident, Key Quality Indicator (KQI) / Key Performance Indicator (KPI), SLA, and Customer Experience Intelligence (CEI) per service per account. For example, intelligence about the customer experience in the Customer Layer allows for a capability to auto-scale in case of bandwidth congestion or for offering self-upgrade service options to customers. Similarly, incidents at the Resource and Infrastructure Layer can enable the OSS system to correct service problems based on some pre-defined policies and using stand-by resources.

Figure 14: Service orchestration optimisation driven by analytics capabilities



The orchestration architecture also presents challenges considering the lack of a well-defined overall framework. The current view based on some of the recommendations seems to be multiple levels of orchestration whereby SDN orchestration, individual or combined NFV orchestration and service orchestration distinctions are being made.

In a longer term, SDN and NFV promise many desirable benefits, including total cost of ownership (TCO) reduction, flexibility (e.g., dynamic lifecycle management and placement of network services), agility (e.g., time-to-market reduction), and network operation intelligence (e.g., automated end-to-end quality-of-service management). The main benefits of NFV and SDN are described in more detail in the remainder of section 4.2.

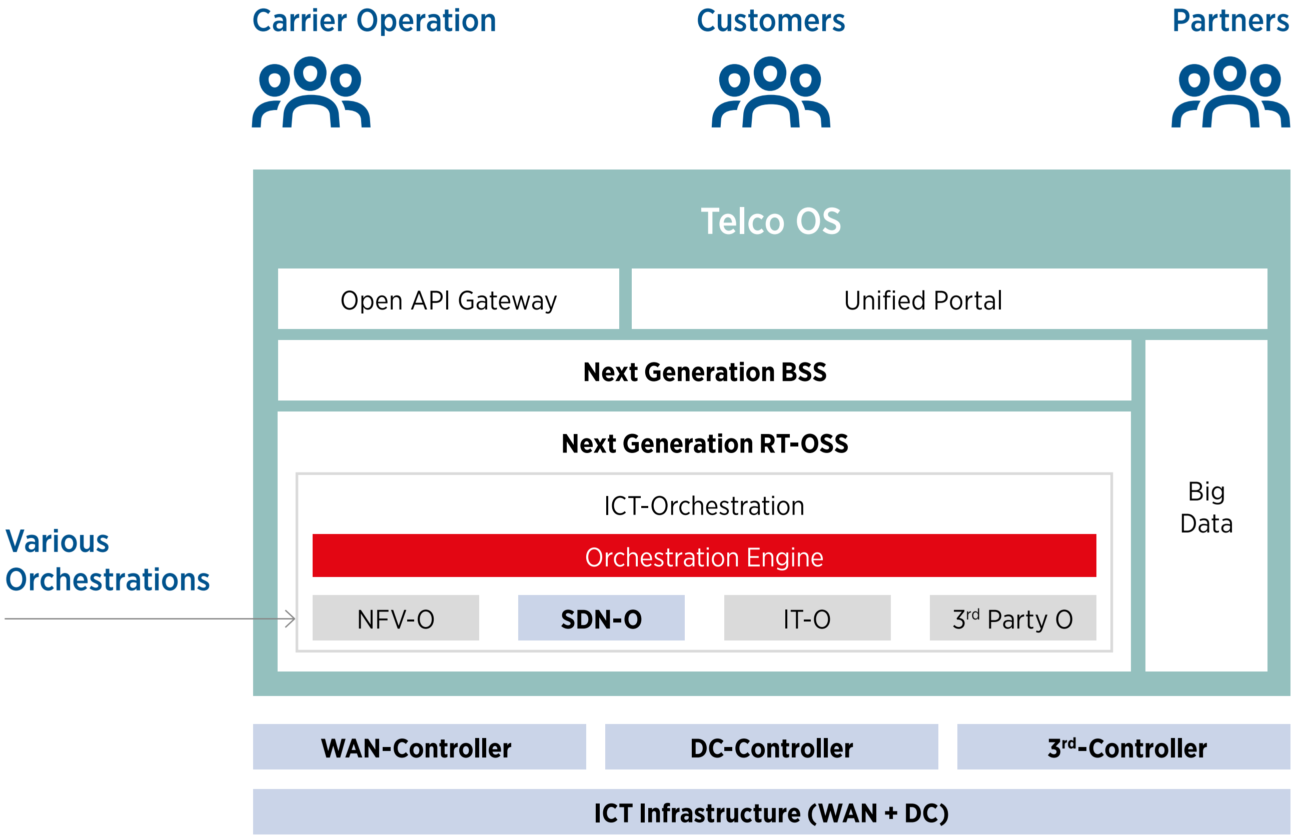
4.2.2.2 Network Slicing

Mobile operator networks currently support a range of vertical services with diverse service requirements. With mobile operators expanding their service portfolio to generate new revenue streams by supporting more vertical services, this diversity is expected to grow even further.

Rather than operators deploying multiple physical networks for each of the vertical services to be supported, one viable approach to satisfy highly diverse service requirements from a single physical infrastructure may be to first virtualize the physical infrastructure with NFV and SDN, and define multiple virtual networks (or network slices) in accordance with the requirements of the vertical services. A network slice can be seen as a dedicated network with its own processing, management and connectivity characteristics that shares the same physical resources with other network slices. Both NFV and SDN are key enabling technologies for overcoming the barriers to the realisation of network slicing.

Using the network slicing concept, it is thus possible to run instances of multiple networks – such as a smart meter network and a public safety communications network – over the same physical infrastructure with possibly different levels of network level security, support for mobility management, latency and throughput.

Figure 15: Orchestration levels



4.2.2.3 Total Cost of Ownership (TCO) Reduction

NFV eliminates hardware dependency of a given network function. This leads to cost savings mainly in four different areas:

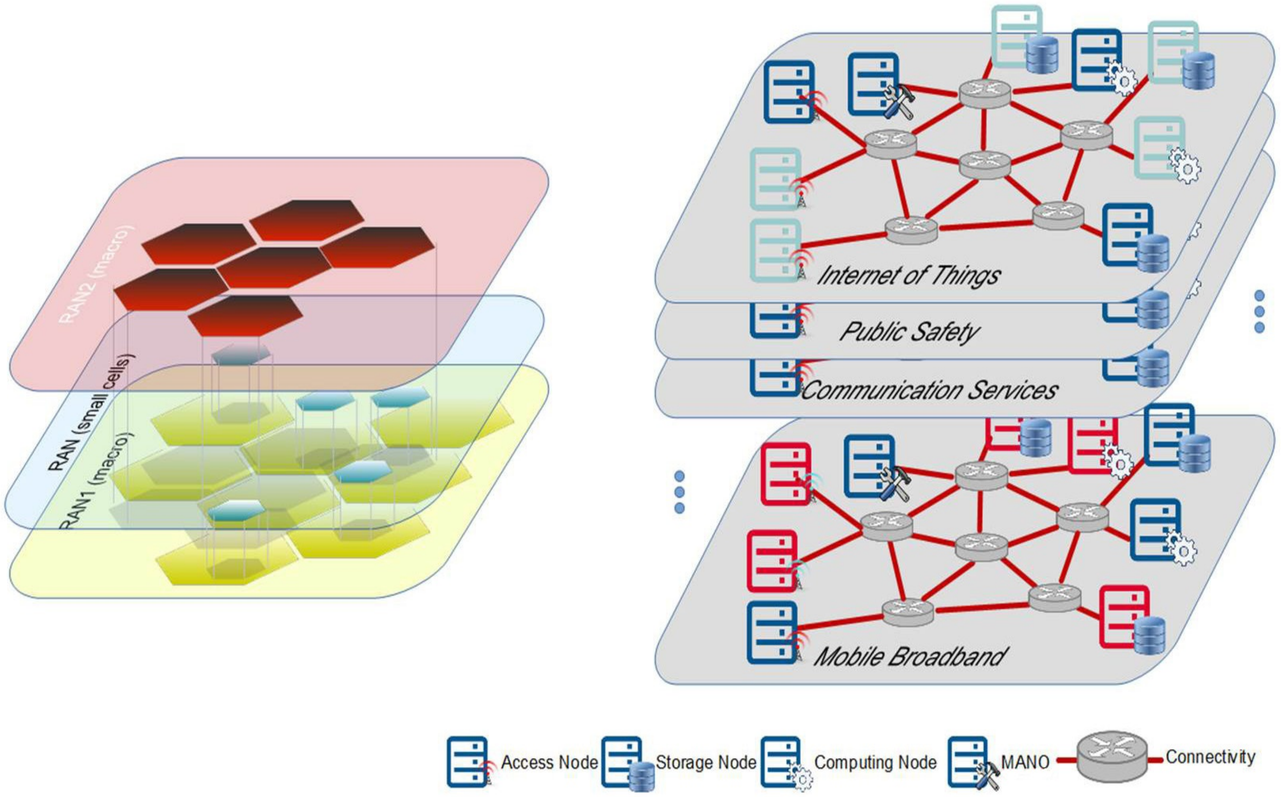
- The servers used to build virtualization infrastructure mostly consist of commercial off-the-shelf (COTS) hardware, which are usually less costly, compared to the previous dedicated and proprietary hardware.
- When additional functions are needed, only the decoupled software functions may be purchased without the costly dedicated hardware. In addition, the individual software can further be cost-optimised (or leaned down in capabilities and configurations).
- The network resources may be provisioned and allocated dynamically and more efficiently during runtime as needed, rather than being statically over-provisioned for the peak usage and not used most of the time.
- Automated lifecycle management and the placement of network services lead to OPEX reduction. This includes energy efficiency savings, as NFV additionally enables consolidating workloads on a smaller number of servers during off-peak hours. This allows putting the other unused servers into energy saving mode or even switch off completely during off-peak hours.

4.2.2.4 Quality of Experience Optimisation

Some performance degradation is likely to happen during the initial commercialisation of NFV and SDN because individual network functions are simply ported and migrated to run on virtualized environments. However, as NFV, SDN, infrastructure and various performance acceleration technologies become mature in 4G, much effort should be spent on understanding and improving the performance of individual network functions to find out how to best run them inside a cloud environment.

In addition to the performance improvements of individual network functions, 5G requires the overall network and infrastructure to provide high performance and guarantee quality of experience from an end-to-end perspective. For example, in the case of certain services, the overall end-to-end delay must not exceed 10 milliseconds (e.g. augmented reality).

Figure 16: Example of network slicing



In order to satisfy and guarantee this end-to-end performance requirement, both the infrastructure and individual network functions used to implement the service must be flexibly and intelligently managed.

One viable approach to reduce the overall end-to-end latency is edge computing (e.g., Mobile Edge Computing) described in the section 4.3, which allows network functions to be flexibly placed at the edge of the network closer to the user. This leads to a reduced end-to-end network path length, and therefore, reduced latency.

4.2.2.5 Robustness

Robustness is one of the most important requirements of mobile networks, and currently various expensive techniques are implemented to achieve a high level of network availability and network robustness. Some of the future mission critical services will emphasise the need for robustness even more.

When using NFV and SDN mechanisms for providing high availability and disaster recovery techniques such as live migration and self-healing in failure scenarios, these must be supported with equivalent or better resilience than the proprietary hardware solutions.

4.3 Mobile Edge Computing

4.3.1 Technology Overview

A close inspection of the 5G opportunities framework reveals that one of the biggest shortcomings of the current 4G network is that it does not lend itself to supporting extreme low latency as well as optimised contents delivery. In order to address this problem and therefore unlock the revenues of this new segment of services, the industry is working towards transforming radio base stations into intelligent service hubs that are capable of delivering highly personalised services directly from the very edge of the network while providing the best possible performance in mobile networks. This architecture is commonly referred to as Mobile Edge Computing (MEC) [21]. Proximity, context, agility and speed can be translated into unique value and thus revenue generation, and can be exploited by operators and application service providers to create a new value chain.

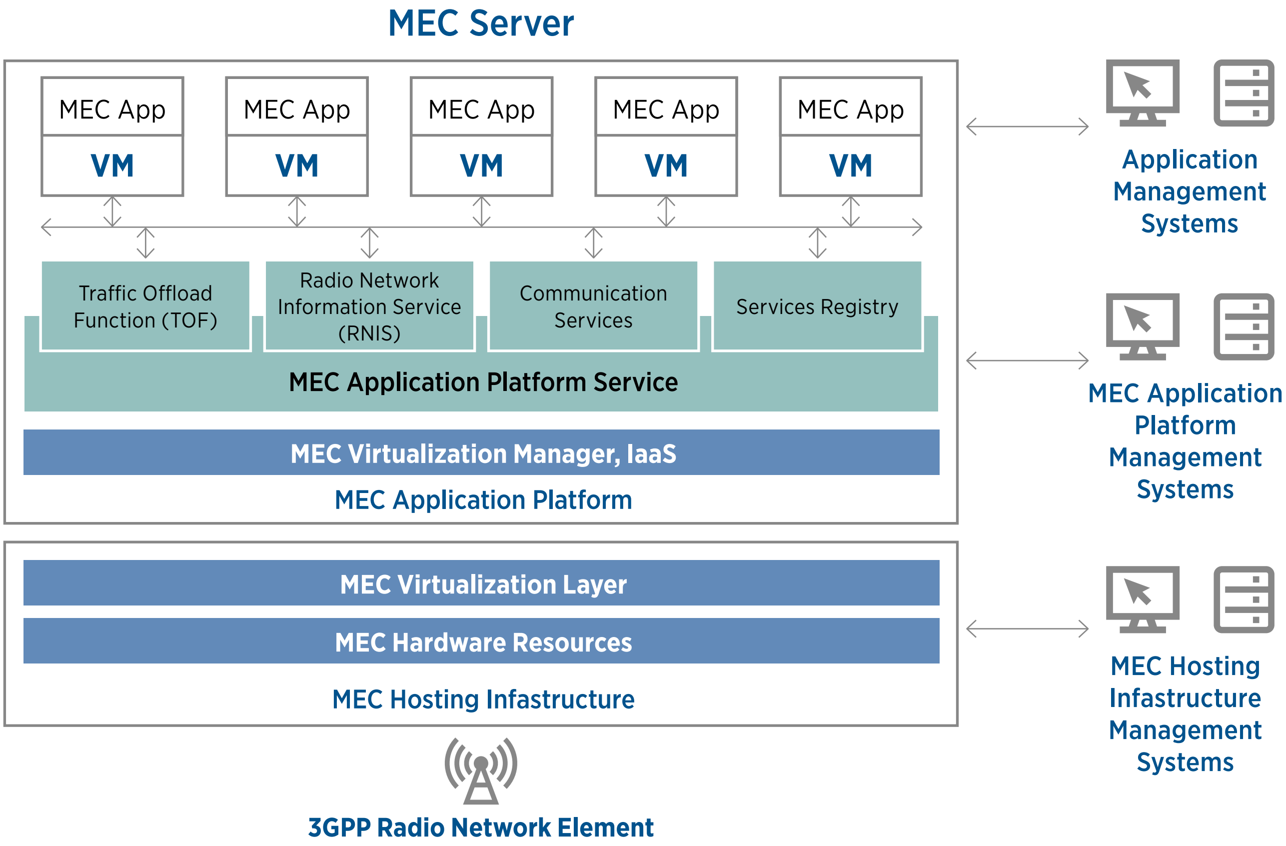
4.3.2 MEC Deployment and Benefits

4.3.2.1 MEC Architecture Highlights

The MEC Platform (Figure 19) consists of the MEC Hosting Infrastructure and the MEC Application Platform.

Source: ETSI ISG MEC

Figure 17: Overview of MEC Server



The MEC Hosting Infrastructure in turn consists of MEC Hardware Resources and the MEC Virtualization Layer providing an abstract interface to applications.

The MEC Application Platform provides the capabilities for hosting applications and consists of:

- The application's Virtualization Manager providing Infrastructure as a Service (IaaS) facilities.
- The Application Platform Services providing middleware application services and infrastructure services to the hosted applications.

The MEC Application Platform offers services to the hosted services and in particular, it offers infrastructure services consisting of a registry of the services that the MEC Server can offer and well-defined APIs to allow the hosted applications to communicate with the application platform provided services. Other functionality of the MEC Application Platform includes:

- Radio Network Information Services (RNIS) that provide authorised applications with low-level radio and network information.
- The Traffic Offload Function (TOF) service that prioritises traffic and routes the selected, policy-based user-data stream to and from applications that are authorised to receive the data.

4.3.2.2 Key Enabler Technologies

In recent years, the telecommunications and IT industries have implemented several technological advancements that will serve as key enablers for the development and success of Mobile Edge Computing solutions. In particular, progress in the area of cloud computing and virtualization will allow applications to be deployed and run on top of the MEC platform in a flexible, efficient and scalable way. Readily available hardware platforms based on highly standardised IT components that can be provided in a competitive manner will be key to the economy of scale of MEC.

4.3.2.3 Technical Challenges

In order to promote and accelerate the advancement of Mobile Edge Computing, the telecommunications community must overcome a diverse set of technical challenges. These challenges are as follows:

- **Network integration**. The introduction of a Mobile Edge Computing server is intended to be transparent to the 3GPP network architecture and the existing interfaces. User Equipment (UE) and core network elements that comply with the existing 3GPP specifications should not be affected by the presence of the MEC server or the applications being hosted on it. The 3GPP protocols and procedures should run and operate without affecting the SLAs.
- **Security**. The MEC platform poses a number of security challenges. The platform needs to fulfil 3GPP-related security requirements while providing a secure sandbox for applications, i.e. isolating applications as much as possible from the burden of having to relate to all the implications of 3GPP security, operator security policies and local regulatory rules.

  Another important aspect to consider is the new physical security constraints arising from deployment scenarios. With regard to the MEC platform deployed at the LTE macro base station, the available physical security is very poor in comparison with that at large data centres. This means that the MEC platform needs to be engineered in a way that will provide protection from both logical intrusions as well as physical ones. The MEC platform therefore needs to establish a trusted computing platform that is resilient to a multitude of attack vectors.

  In addition to trusting the platform itself, operators will be concerned about security implications with regard to the deployment of third-party software applications. In the particular case of Mobile Edge Computing, this concern may be exacerbated by the fact that these third-party software applications are deployed in very close proximity to their most valued asset, the Radio Access Network. The platform needs to provide the mechanisms to ensure that the virtual machines which contain the packaged application software (destined for deployment) come from a trusted source, are authenticated and authorised and are, therefore, safe to be incorporated in the platform.

- **Performance**. Since the Mobile Edge Computing platform provides a virtualized environment where software applications are hosted, the main challenge facing the industry concerns the problem of extracting maximum performance while minimising the impact of virtualization. The “extra” layers introduced by virtualization should not degrade performance, particularly with regard to applications that require the intensive use of hardware resources (e.g. intensive input/output (I/O) or computing) or require low latency.

4.3.2.4 Example of MEC Application

To illustrate the advantages of MEC, we will consider the following two examples of the application of MEC: mobile content delivery acceleration and RAN-aware content optimisation.

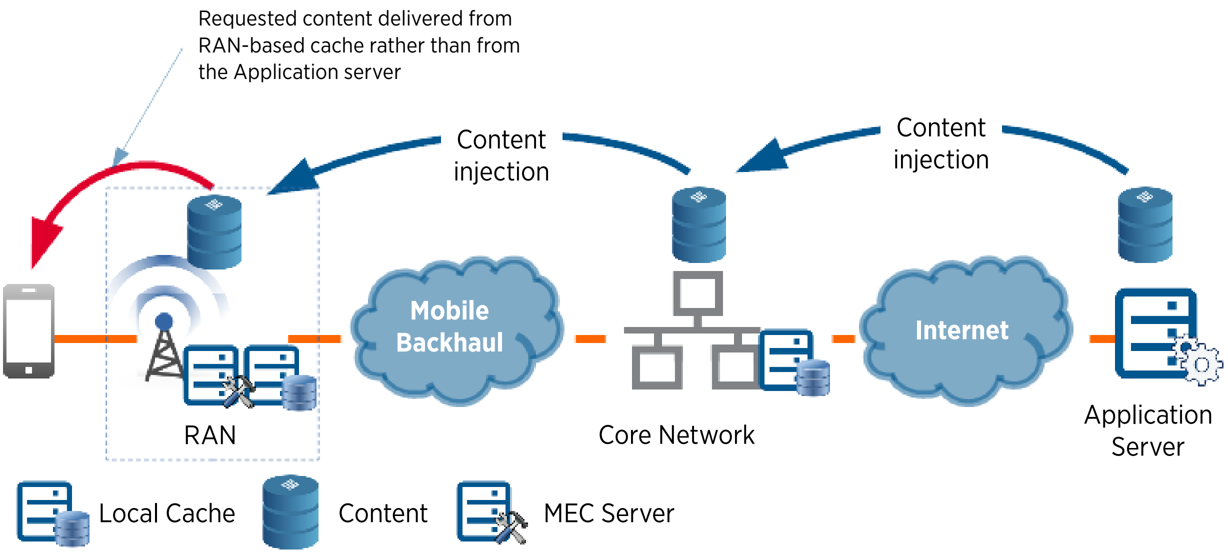
Content that is requested frequently is cached in the RAN. The UE gets the service from the base station directly, which decreases the delay between the UE and the application server and relieves bandwidth tension on the backhaul.

The RAN-aware application deployed in the MEC server exposes accurate cell and subscriber radio interface information (cell load, link quality) to the content optimiser, enabling dynamic content optimisation, improving QoE, network efficiency as well as new service and revenue opportunities. Dynamic content optimisation enhances video delivery through reduced stalling and reduced time-to-market.

Besides the two examples provided, MEC can also be used to support advanced services requiring very low latency (e.g. gaming, augmented reality) as well as services that are local (e.g. communications within an enterprise).

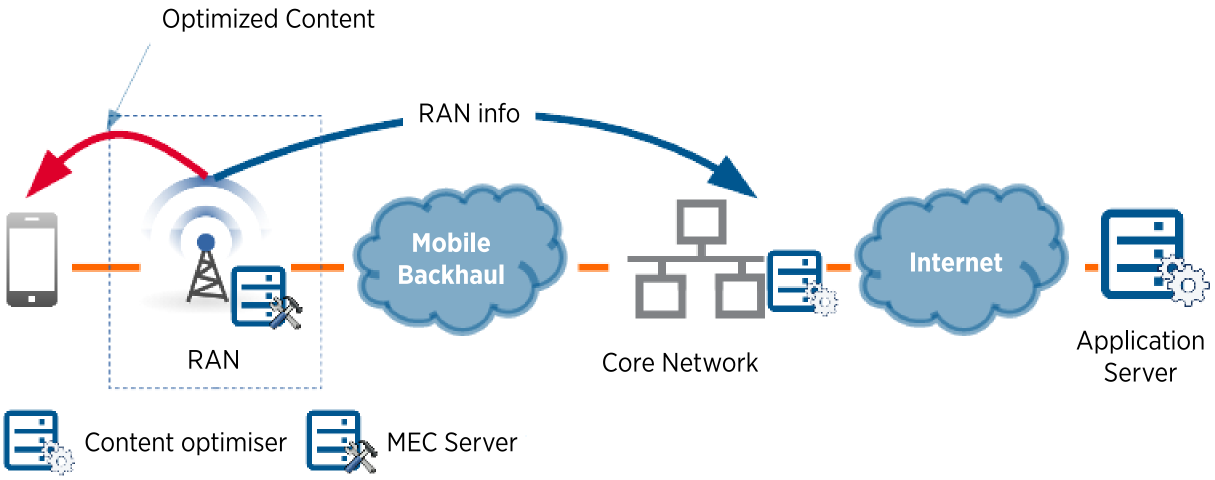
Source: ETSI ISG MEC, recreated by the GSMA

Figure 18: Content delivery acceleration



Source: ETSI ISG MEC, recreated by the GSMA

Figure 19: RAN-aware content optimisation



4.4 Internet of Things

4.4.1 Technology Overview

The Internet of Things (IoT) has been identified as one of the major opportunities in 5G, especially in the cases where the mobile operator is in the position of attacking the value chain and increasing the share of end-to-end service revenues. It is worth remembering that IoT is a generic term for describing those applications where devices communicate with each other or with a server. IoT applications therefore vary widely in terms of data throughput rates, latency, communication frequency, costs and power consumption requirements. As a consequence, there cannot be a onesizefitsall solution for IoT and in fact the technology advancements to better support this type of devices has progressed on many fronts. The powerful combination of these advancements and the capability to create dedicated network slices to fulfil the needs of each type of IoT service has the potential of placing mobile operators in the position of exploiting this new revenue stream.

4.4.2 IoT Deployment and Benefits

4.4.2.1 Technical Evolution for Supporting IoT

This section introduces the state of the art in terms of capabilities of mobile networks and then gives an indication on how such capabilities can support the 5G use cases.

Starting from their Release 10, 3GPP defined functionality in their technical specifications aimed to address requirements specific to the support of IoT (Machine Type Communications (MTC) in 3GPP parlance). These improvements have been evolved and refined over the course of subsequent 3GPP releases, while other functionality initially earmarked for MTC has been later generalised to make it applicable to other types of devices too.

4.4.2.2 Signalling Congestion Overload

It was envisaged that there may be services where a very large amount of devices transmit information to the network in the same timeframe (e.g. electricity metres reporting daily power consumption at 4:00). This situation is likely to cause signalling overload and – depending on the amount of data to be transmitted – congestion, too. 3GPP therefore introduced enhancements in Release 10, such as back-off timers, aimed to protect the operator network.

For detailed information on how some of the 3GPP features can be utilised when designing M2M/IoT applications, refer to GSMA publication "CLP.03 IoT Device Connection Efficiency Guidelines" [16].

4.4.2.3 Power Saving Mode

From Release 12 onwards, more efforts have gone into addressing the need to support devices capable of operating for 10 years on a 5Wh battery. This resulted in the specification of Power Saving Mode (PSM), which is applicable to all 3GPP radio access technologies (GSM / Wideband Code Division Multiple Access (WCDMA) / LTE).

A Power Saving Mode solution works at Non Access Stratum level by allowing the UE to negotiate with the network periods during which the UE stops all access stratum activities. During these periods, the UE is not reachable.

In situations where the device needs to be reached at unscheduled times or when the data is not delay tolerant, PSM loses its advantages since the UE needs to frequently reconnect with the network, increasing battery consumption dramatically. A more flexible approach that addresses the scenarios not suitable for PSM is to enhance Discontinuous Reception (DRX) operation. Unlike PSM, DRX can make the UE reachable during pre-defined occasions without resulting in unnecessary signalling. In addition, unlike PSM, the extended DRX solution is primarily an access stratum solution.

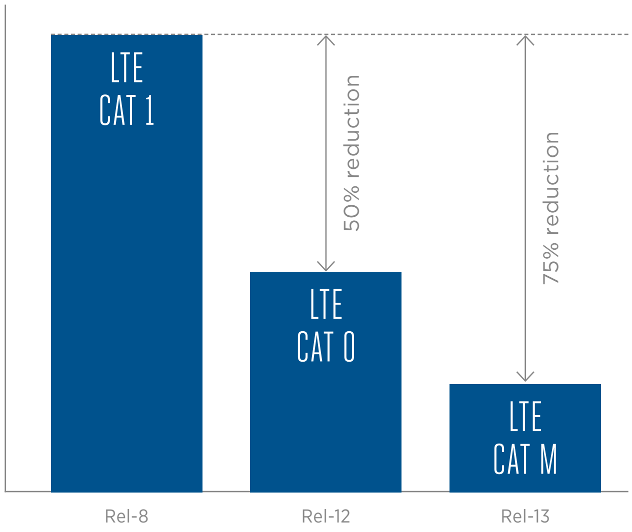
4.4.2.4 Radio Enhancements for Low Complexity Devices in Challenging Coverage Conditions

As in many cases, a device designed for MTC only needs a subset of the capabilities of a consumer oriented device. 3GPP specified a new UE category, Cat 0. This new category enables 50% lower modem complexity compared to a single-band single-RAT (Radio Access Technology) LTE category 1 (Cat 1) modem through a reduction of the highest supported data rate to 1 Mbps, something considered more than suitable for a vast subset of IoT applications. A Cat 0 UE will operate with a single receive antenna, and half-duplex frequency-division duplexing (FDD), meaning that transmission and reception are not performed simultaneously.

A new 3GPP work item "Further LTE Physical Layer Enhancements for MTC" (not yet completed at the time of writing) aims to define a new UE (referred to as Cat M or Cat -1), which reduces complexity by about a further 25% compared to LTE UE Cat 0, and also improves coverage.

In addition to reducing complexity, 3GPP also worked on enhancing the reachability of the UEs designed for MTC use. The sensitivity of the device may be improved by 15 to 20 dB, making MTC devices reachable in places where there would normally be no service.

Figure 20: LTE UE category complexity reduction forecast



4.4.2.5 Ultra-Low Complexity and Low Throughput IoT

A study item produced by 3GPP GERAN (3GPP TR 45.820 [17]) has investigated the support for Ultra-Low Complexity and Low Throughput Internet of Things. The aim is to support the following capabilities:

- Provide a data rate of at least 160 bps on both uplink and downlink, which should be achieved in an extended coverage of 20 dB compared to legacy GPRS.
- Scale to support a massive number of MTC mobile stations. This is the same traffic model from 3GPP TR 36.888 "Study on provision of low-cost Machine Type Communications (MTC) User Equipment's (UEs) based on LTE" [18] with the modification to use 40 low throughput devices (rather than 3 smart meters) per home.
- Reduce power consumption of MTC mobile stations compared with legacy GPRS, for reaching up to ten years battery life with a battery capacity of 5 Watt-hours. The requirement should be fulfilled in challenging coverage conditions (e.g. 20 dB extension) too.
- Maximise the reduction in complexity of Mobile Termination (MT) compared to that of legacy GPRS MT.
- Avoid negative impacts to legacy GSM/WCDMA/LTE system(s) deployed in the same frequency band and adhere to the regulatory requirements applying to the spectrum bands in which the system operates.
- Minimise impacts to GPRS/EDGE base station hardware.
- Identify Core Network Architecture, security framework and Radio Access Network-Core Network interface (e.g. S1 or Gb), and associated protocol stacks, suitable for the M2M market in the 2017 and onwards timeframe. Restrict use to a simple QoS model.

The study envisages that a new radio access technology as well as an optimised core network may be defined as a result of this study. At the time of publishing of this report, work is ongoing in 3GPP working groups in this area:

- the architecture group (SA2) has completed a study on architecture enhancements for Narrow Band IoT (NB-IoT);
- the Radio Access Network group has approved a new work item on Narrow Band IoT solutions;
- The GSM/EDGE RAN (GERAN) group has approved a new work item on the Extended-Coverage GSM (EC-GSM).

The aim is to complete the normative work on these activities in Release 13.

4.5 Evolved Broadcast Capabilities

4.5.1 Technology Overview

Broadcast services have been earmarked as a potential area of growth in the future (see section 3.5.2). While it would be premature to speculate on whether the new 5G radio access technology will provide wide area coverage, one enabler of advanced broadcast services that is available in 4G networks is multicast services using point to multipoint transmission either over multiple cells such as the evolved Multimedia Broadcast / Multicast Service (eMBMS) or over single cell such as the single cell point to multipoint (SC-PTM). By providing greater bandwidth capability and more efficient spectrum usage, eMBMS and SC-PTM are more apt to meet the increasing demands for bandwidth intensive broadcast/multicast services.

Another use case enabled by the advanced capabilities of eMBMS/SC-PTM is support for multimedia mission critical communications.

4.5.2 Broadcast Technologies Deployment and Benefits

4.5.2.1 Evolved Multimedia Broadcast / Multicast Service

The evolved Multimedia Broadcast / Multicast Service is the LTE evolution of Multimedia Broadcast / Multicast Service (MBMS) which was introduced for WCDMA in 3GPP Release 6 and described in 3GPP TS 23.246 [9] and 3GPP TS 25.346 [10]. eMBMS was introduced in Release 9.

eMBMS allows simultaneous transmission of identical information/content (e.g. a news stream) to multiple terminals typically dispersed over a large area and across a number of different cells.

eMBMS supports both broadcast and multicast services, which are different services though closely related. For broadcast services, there is no tracking of the user's movement and the user can receive the content without notifying the network. Multicast services require a user to join a multicast group and the user's movements are tracked in order to match radio resources to the number of users in a given cell. It should also be noted that eMBMS multicast may be configured to use point-to-point transmission instead of point-to-multipoint transmission where a cell is sparsely populated by users of a given multicast group.

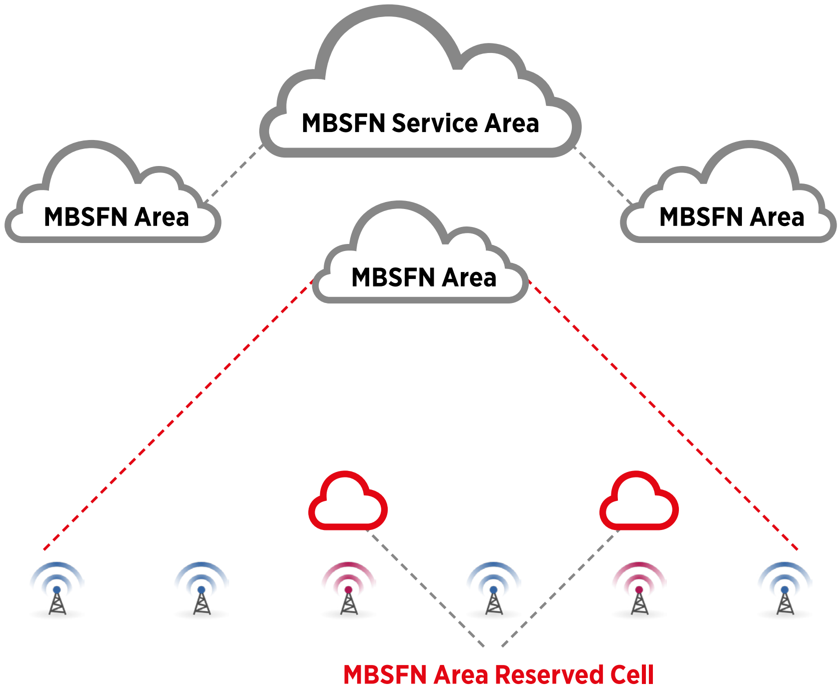
eMBMS provides a substantial saving of network radio resources by replacing several point- to-point transmissions of the same content with a single broadcast cell-wide transmission. The radio resource saving (especially high in urban centres or a full sports stadium) is partially offset by the need to dimension broadcast transmissions for the worst case scenario, where a target terminal is located at the edge of a given cell. In addition, there is a limited signal-to-noise ratio that may be achieved within a cell, and in particular at the cell edge, which limits the data rates that can be achieved and this is especially true in the case of large cells. One technique used to increase the efficiency of MBMS at cell edge is to exploit the fact that the UE is likely to receive and then combine the same information from multiple cells. Such combining of different transmissions from multiple cells is known as macro diversity.

To provide eMBMS bearer services, a new specific functional entity – the Broadcast Multicast Service Centre (BM-SC) – was introduced for MBMS. The BM-SC provides functions for eMBMS user service provisioning and the delivery of eMBMS transmissions. It may also serve as an entry point into the public land mobile network (PLMN) for content provider eMBMS transmissions.

The LTE eMBMS introduces several enhancements to the basic 3G MBMS service.

Source: 3GPP, recreated by GSMA

Figure 21: eMBMS hierarchical structure



For the purpose of delivering eMBMS services, the network can be subdivided into the following logical areas:

- **MBSFN Synchronisation Area:** An area of the network where all eNodeBs can be synchronised and perform multicast-broadcast single-frequency network (MBSFN) transmissions. MBSFN Synchronisation Areas are capable of supporting one or more MBSFN Areas. On a given frequency layer, an eNodeB can only belong to one MBSFN Synchronisation Area. MBSFN Synchronisation Areas are independent from the definition of MBMS Service Areas.
- **MBSFN Area:** An MBSFN Area consists of a group of cells within an MBSFN Synchronisation Area of a network, which are co-ordinated to achieve an MBSFN Transmission (i.e. a technique whereby identical waveforms are transmitted at the same time from multiple cells – which is seen by the UE as a single transmission). All cells within an MBSFN Area (apart from reserved cells) contribute to the MBSFN Transmission and advertise its availability.
- **MBSFN Area Reserved Cell:** A cell within a MBSFN Area that does not contribute to the MBSFN Transmission. The cell may be allowed to transmit for other services.

The core network architecture was also enriched with two further logical nodes, the Multi-cell/multicast Coordination Entity (MCE) and the eMBMS Gateway (eMBMS GW).

The MCE serves an eNodeB (eNB) and performs:

- The admission control and the allocation of the radio resources used by all eNBs in the MBSFN area for multi-cell MBMS transmissions using MBSFN operation.
- Counting and acquisition of counting results for MBMS service(s).
- Resumption/suspension of MBMS session(s) within MBSFN area(s) based on e.g. the Allocation and Retention Priority (ARP) and/or the counting results for the corresponding MBMS service(s).

The MBMS GW principal function is the sending/broadcasting of MBMS packets to each eNB transmitting the service. The MBMS GW uses IP Multicast as the means of forwarding MBMS user data to the eNBs.

Compared to MBMS, eMBMS enables increased flexibility for spectrum usage via the capability to reserve network resources for multicasting for a duration of a session and to release those resources at the end of a session. In addition, eMBMS can also operate across a group of cells (MBSFN Area) so that reception at the UE remains consistent as the UE moves between cells. This is due to a frequency block being reserved across the group of cells and transmissions synchronised between the cells, which results in the UE picking up the combined signal with no interference. In addition, based on user distribution, some cells may be precluded from the MBSFN area. The inherent greater bandwidth available in LTE also permits better quality video to be multicast than was the case for MBMS.

4.5.2.2 Single Cell Point to Multipoint

Single Cell Point To Multipoint (SC-PTM) is the eMBMS evolution introduced in Release 13. SC-PTM keeps the eMBMS system architecture while enhancements are specified in the air interface to improve the spectrum efficiency as well as to reduce the latency. SC-PTM is especially suitable for scenarios where broadcast service is expected to be delivered to a limited number of cells due to user interests and the concerned cells may dynamically change due to user movement. SC-PTM can optimize the use of network resources for the provision of bandwidth intensive broadcast services.

With SC-PTM, a list of cell identities where a broadcast service should be transmitted will be provided to E-UTRAN during eMBMS session establishment and SC-PTM transmission will be performed on those cells. User's movements are tracked so that the network knows user distributions for each multicast group and can then decide the list of cell identities (e.g. the cells with users located).

SC-PTM transfers the broadcast service using the existing channel for unicast transmission. The SC-PTM scheduling is quite agile and radio resources could be dynamically assigned in time and frequency domain based on real time traffic load. This enables the the LTE system to support a varied multimedia content in a dynamic fashion based on user demand. Both SC-PTM and unicast have the same radio frame structure, and the radio resources could be flexibly shared between them even in the same radio subframe, which consequently means that the spectrum resources can be fully utilized without any waste. Multiple antenna port transmission also comes for free for SC-PTM, which will significantly increase the spectrum efficiency and robustness.

For SC-PTM, synchronization between multiple sites is not required, and this will significantly simplify the network planning and deployment thus saves CAPEX/OPEX. In case of critical situations (e.g. natural disaster or emergency rescue), radio sites could be fast deployed to provide support of SC-PTM, to cope with the increasing communication demand in critical communications.

4.6 Proximity Services

4.6.1 Technology Overview

Proximity services as specified in Release 13 of 3GPP already offer an invaluable tool for public safety communities, however they can also be considered enablers for some of the broadcast use cases and will play a role in the vehicle-to-everything (V2X) area. They may also be considered enablers for other Professional Mobile Radio (PMR) type services, such as mission critical communications for large enterprises in industries such as power and transport.

4.6.2 Proximity Services Deployment and Benefits

4.6.2.1 Proximity Services Description

Proximity Services (ProSe), also known as Device to Device (D2D) proximity services in 3GPP RAN, are described in 3GPP TS 22.278 [15] and consist broadly of two main service classes:

- **ProSe Discovery:** ProSe discovery allows a device (UE A) to detect the presence of another device nearby (UE B) provided that UE A has expressed interest in discovering and has been granted the permission to discover UE B.
- **ProSe Communication:** This service enables two devices located in proximity of each other to establish a direct communication without network interaction. A simplified version of ProSe Communication relies on one-to-many communication whereby a device broadcasts its message (walkie-talkie like).

Release 12 Stage 3 specifications focused mainly on:

- One-to-many LTE D2D communication for public safety applications inside and outside the coverage (intra-cell and inter-cell).
- In-coverage LTE D2D discovery in particular for commercial ProSe services (intra-cell and inter-cell).
- Multi-PLMN and Multi-Carrier support for in-coverage LTE D2D discovery.
- Use of exclusively licensed spectrum for such LTE D2D operation.

In order to enable LTE D2D operation, 3GPP defined what is called a “sidelink”. It corresponds to a new UE-to-UE interface, called PC5 interface, which conveys ProSe direct communication and ProSe Direct Discovery.

The application layer for ProSe services is not specified in Release 12.

3GPP is currently working on Release 13 enhancements for LTE D2D operation focusing on:

- Specification of discovery for partial and outside network coverage for public safety.
- Introduction of UE-to-Network Relays for coverage extension.
- Prioritisation of different groups of users.
- D2D enhancements for Multi-Carrier and Multi-PLMN support.
- Mission Critical Push To Talk support.

The network broadcasts or provides in a dedicated channel the radio resources dedicated to ProSe. For a ProSe discovery event, the announcing UE receives from the network the code that can be broadcast (ProSe code), the time for which this code is valid, and the allowed repetition time. As the radio resources for ProSe are shared and not managed, the network needs to ensure that not too many UEs try to announce at once. The discoverer’s UE will receive from the network a filter that allows it to determine which of the monitored ProSe codes correspond to UEs of interest. When the monitoring UE matches a ProSe code, it reports this event to the network.

The mobile operator retains control of both the frequencies allowed for ProSe communication and discovery as well as the permission to use the service.

ProSe Discovery also provides additional modes of operation, for example whereby a user requests if a certain service or person is in proximity.

4.6.2.2 ProSe Communication

Due to the limited range, coexistence issues with legacy Wide Area Network (WAN) communications and its complexity, the ProSe Communication primary use case is for public safety authorities that require operation in the absence of network coverage. However, enhancements to make ProSe Communication applicable to vehicle-to-everything (V2X) communications are also being studied in the framework of Release 13 and Release 14.

4.6.2.3 ProSe Discovery

ProSe Discovery has both commercial and public safety applicability. ProSe Discovery can be further subdivided into two modes of operation, “open” and “restricted” discovery. When open discovery mode is used, any ProSe enabled device can discover the device advertising its presence. Open discovery is therefore suitable for scenarios where, for example, a shop advertises its presence to subscribers who declared interest in receiving information from that shop or in the category of service that the shop provides. When restricted discovery mode is used, UE A’s announcement of its presence can only be successfully received by those UEs that have been authorised. A typical example of restricted discovery is social network applications where the announcement of a user can only be deciphered by the user’s buddies/friends on that particular social network.

4.6.2.4 ProSe for V2X

From RAN point of view, the V2X communication will be supported based on at least the PC5 interface. The carrier of V2X over PC5 could either be a sharing carrier, i.e. Uu and PC5 share the carrier, or dedicated carrier only for V2X communication, i.e. different carriers for Uu and PC5. For both sharing and dedicated carrier deployment scenarios, the resource allocation should be under the operator network control when V2X communication peer nodes are in the network coverage. Semi-persistence resource allocation or UE-assistant resource selection mechanism are also considered according to the V2X service characteristics, e.g. periodicity, message size and communication range and etc., for reducing the network load in future study.

Another way for supporting V2X is to implement the communication through the Uu interface. In this scenario, one V2X device may transmit the V2X messages to the network nodes through the Uu interface in uplink. The network nodes may play a role as Infrastructure Unit for receiving V2X messages. If the message needs to be transmitted to other V2X node, e.g. V2V communication through Uu, the network nodes could broadcast the V2X message received to the other vehicles through the Uu in downlink by means of for example eMBMS

4.6.2.5 Roadblocks to Proximity Services

The commercial deployment of ProSe open and restricted discovery will require additional work in order to ensure coordinated approach both at lower network layers (e.g. spectrum allocation, multi-operator operations) and application layers (e.g. business type categorisation).

4.7 Traffic Management and Quality of Service

4.7.1 Technology Overview

The categories of services that will be deployed in a 5G system will be characterised by a broad spectrum of Quality of Service and priority: traffic management will therefore be a key component in enabling operators to offer these services with the required quality of experience as well as to ensure that the network resources are used efficiently. Furthermore, by exposing QoS capability to third-party service providers, mobile operators have a powerful tool to offer to fulfil the demands of vertical industry services. Ultimately, the availability of mechanisms to control the delivered quality transforms QoS into a marketable asset.

4.7.2 Traffic Management Deployment and Benefits

4.7.2.1 Rationale for Traffic Management and QoS Management

In 2015, mobile operators face a 54% year-on-year growth of global average data traffic. With the rapid development of wireless communication and the IoT industry, mobile network connections will increase to several tens of billions.

This dramatic growth of data traffic has made the capability of a mobile operator to manage the network traffic in order to provide Quality of Service a critical factor affecting operators’ network development plans as well as customer experience.

Although 4G and 5G are based on the Internet Protocol, which has a number of deficiencies, a significant amount of work has been carried out to develop traffic management techniques as well as to enable the exposure of network capabilities to third parties.

4.7.2.2 Exposed QoS Capabilities

End-to-end Quality of Service is primarily determined by end-to-end (E2E) delay, delay variation (“jitter”) and throughput. However, effective traffic management techniques are also applied taking into consideration a number of other attributes such as traffic types, relative priorities, source and destination addresses.

The 4G LTE network offers application-level QoS differentiation and traffic management in an architecture that incorporates mechanisms to control bandwidth, latency, potentially jitter and packet loss. This architecture is scalable for end-to-end QoS by the controlling mechanisms at an aggregate level. Application traffic can be categorised into multiple types, and put into corresponding classes of services defined with QoS parameters. This is necessary because different types of traffic are more sensitive to certain parameters than others. For example, while the quality of a video is primarily determined by the throughput, the quality of a VoIP session depends on the jitter.

4G LTE supports end-to-end QoS. 3GPP TS 23.203 [19] defines a number of parameters associated with the connection between a mobile device and the network, which consists of a radio access bearer (RAB) and a core network bearer (packet data network (PDN) connection):

- QoS Class Identifier (QCI). Each QCI value is defined by Resource Type, Priority, Packet Delay Budget and Packet Error Loss Rate.
- Allocation and Retention Priority (ARP).
- Maximum Bit Rate (MBR).
- Guaranteed Bit Rate (GBR).
- Aggregate Maximum Bit Rate (AMBR).

3GPP has standardised a number of QCI values. This allows, for example, a mobile operator to request another 3GPP network to apply a pre-determined Quality of Service to a bearer. For example, successful establishment bearers for a VoLTE call while roaming is based on the use of QCI values.

The architecture of the entire capability exposure system (Figure 24) consists of three levels: Network Layer, Capability Layer and Application Layer. The Application Layer and the Capability Layer form the Capability Exposure Platform.

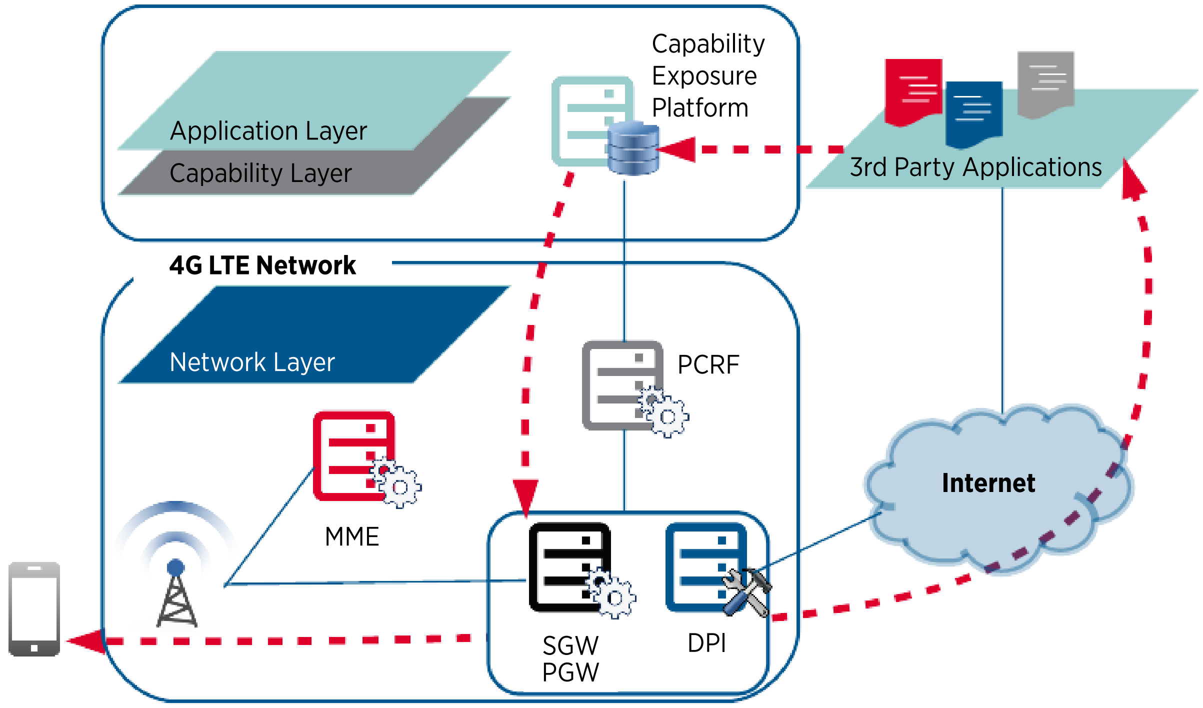
The Network Layer is the network infrastructure, including RAN, EPC, and other systems of the mobile network. The network layer is the foundation of network capability exposure, it implements the policy delivered by the capability layer, including resource allocation, congestion control, and QoS policies.

The Capability Layer is a platform for network capability exposure located between the network layer and application layer. The functions in this layer include network information analysis and abstraction, network capability orchestration, capability exposure services processing, and providing a specific API for the application layer.

The Application Layer can invoke the internal API in the mobile network provided by the capability layer, and provide an external API for third-party service providers. The third-party API management module in the application layer is responsible for the authentication, charging and security of the API, as well as other management tasks.

Source: 3GPP

Figure 22: Capability exposure reference architecture



4.7.2.3 Examples of Usage of QoS and Traffic Management

This section describes how traffic management can be applied to provide advanced services. A number of examples follow:

- **Customized QoS Assurance on VIP CP/SP:** Service providers can offer prioritised QoS rankings to selected customers to guarantee their preferential treatment across the operator's network. The customisation could be applied to one or more of the QoS parameters, thus improving the experience for services that demand, for example, very low latency or minimum bandwidth.
- **Automatic QoS Assurance on VIP CP/SP:** An automatic QoS assurance can be given to the network operator's third-party partners of strategic value, or to the network operator's own services.
- **One Click to Enhance Bandwidth:** Subscribers are given the possibility to ask for a higher data rate ("Bandwidth on Demand").
- **Bandwidth Limitation and Recovery:** The network can implement a protection mechanism for subscribers at risk of running out of their data allowance by, for example, reducing the maximum bandwidth allowed rather than disconnecting the user, thus allowing for a better experience.

4.8 IP Multimedia Subsystem

4.8.1 Technology Overview

While the main areas of revenue growth opportunities in 5G will derive from new market sectors, as discussed in section 3.3.3, legacy telecommunication services still play a role in the operator's offer portfolio.
With the IP Multimedia Subsystem (IMS) being the service platform of choice in 3GPP, all the traditional telecommunication services, including supplementary services, will be delivered via an application server through IMS. Moreover, IMS is not limited to emulating legacy services and in fact several services, defined both in 3GPP and outside, are exploiting the IMS capabilities. Examples include Rich Communication Services (RCS), push to talk (for public safety) and WebRTC interworking.

4.8.2 IMS Deployment and Benefits

4.8.2.1 IMS Development and Evolution

Historically, on legacy 2G/3G wireless access, voice/SMS services and data services were cleanly separated into the circuit switched (CS) world of Mobile Switching Centres (MSCs) and the packet switched (PS) world of GPRS, respectively. This situation changed with the advent of Evolved Universal Terrestrial Radio Access Network (E-UTRAN) or Long Term Evolution (LTE) as LTE provides only PS access with no support of CS technology and thus no native support of voice/SMS. Whilst interim solutions were used for the provision of voice (i.e. Circuit Switched FallBack (CSFB) via legacy RATs) and SMS (i.e. SMS over SGs via MSC), the industry agreed evolution is to provide voice and SMS via the all-IP PS network and utilising the 3GPP IP Multimedia Subsystem as the service layer.

From its inception, IMS has been intended to be a platform that would enable service innovation in the operator community. IMS enables operators to:

- Leverage the strengths they have, such as the large amount of existing users, wide range of roaming agreements, interconnecting and interoperating capabilities.
- Provide a framework for exposing their network capabilities to third-party developers and allow faster and wider service development.

4.8.2.2 Common Services Infrastructure

The IMS provides a single, Session Initiation Protocol (SIP) based session control that is common to all services and applications. This means that once an IMS is deployed, new services can be added by installing application servers for those services. While the service logic is obviously specific to the service, IMS consolidates the service layer functionality leading to cost savings and economy of scale. Users gain access to their full set of services via a single IMS registration. When a service is invoked, the IMS takes care of invoking the relevant application server.

4.8.2.3 Access Agnostic Infrastructure

IMS has been designed from its inception as a gateway between the application layer and the transport layer; as a consequence, IMS is largely independent of the underlying core and access network. Testament to the success of this design is that IMS is successfully deployed both in fixed and mobile networks.

Generic procedures are defined to allow the IMS to interact with the underlying network. In the case of 3GPP mobile systems and in particular the Evolved Packet Core (EPC), the interaction happens via the Policy and Charging Control layer (more specifically through the Policy and Charging Rules Function (PCRF)).

When IMS is deployed, it can deliver services to devices connected via a wide range of access technologies. For example, the same application server is used to deliver voice over LTE and voice over Wi-Fi.

4.8.2.4 Providing Quality of Service

In combination with the Policy Control Function and the underlying packet core, the IMS provides for the required QoS appropriate for its services. For example, in the case of Multimedia Telephony, the IMS provides the same (or better) user experience for voice calls that subscribers experience on the CS network and thus whether the call is a CS call or an IMS controlled PS call is transparent to the end user.

The EPC's Policy Control Function acts as a policy decision point for policy and charging control of all service data flows/applications and IP bearer resources. This function receives resource requests from the IMS layer related to calls/sessions and interacts with the EPC layer to satisfy those requests. This function thus shields the IMS from the underlying packet core network. This function also acts as an information exchange relay between the IMS service layer and the EPC bearer layer to forward event triggers (e.g. inform the IMS that a bearer on which a call is dependent has failed).

4.8.2.5 Service Continuity

Operators can benefit from the IMS Centralized Services (ICS) work performed in 3GPP allowing IMS to also take control of services delivered by legacy circuit switched networks. ICS requires upgrading of some of the legacy nodes (MSC), but when deployed, results in a single service domain being used across all the operator's available accesses.

4.8.2.6 Common NNI

The Network-Network Interface (NNI) is the interconnecting interface between two operator networks. This interface is defined tightly within 3GPP standards (see 3GPP TS 29.165 [11]) and is further profiled within GSMA PRD IR.95 [12] with as little flexibility as possible so that the implementation of one network's service and that of another network requires minimal or no interworking functionality to be placed between them.

The NNI for an IMS based service adheres to QoS principles for the allocation of both signalling and media traffic to an appropriate DiffServ Class of Service.

The inter-IMS (NNI) interface is a generic NNI that supports all of the IMS-based services and is independent of the specific access technology used by the UE. The ability of the NNI to support all of the IMS-based services results in a consolidation of interconnect points for a given operator, regardless of whether this is done bilaterally or via an Internetwork Packet Exchange (IPX).

4.8.2.7 IMS Providing Roaming

The roaming interface is also made up of a signalling plane element and a media plane element. In addition, as operators move forwards with the introduction of VoLTE and the Evolved Packet Core, Policy Information may be exchanged between the visited and home networks.

A roaming architecture and interfaces have been defined so that the service that is supported when the customer is in their home network is available to them in the same fashion when they roam to a visited network (i.e. Local Breakout for VoLTE). This in turn means that QoS characteristics can be maintained and chargeable events can be communicated from visited to home networks as needed.

The architecture provides support for Lawful Interception and Emergency Calls for voice services whilst in the visited network.

4.8.2.8 IMS Providing Service Capability Exposure

Operator deployments of IMS-based services enable interaction between operators' customers using the interoperability and interconnect instilled in the definitions of those services. However, rather than simply deploying IMS services as services in their own right, there is likely to be considerable value in exposing these service as "capabilities" to third party developers.

IMS network service exposure contains three dimensions:

- IMS basic service capabilities, which includes Capability Exchange, Network Address Book, Real-time Voice and Video Telephony plus other enriched call services (e.g. Chat, File Transfer, etc.).
- Inter-working and interoperability between different network technologies. The IMS service platform shall support different mobile and fixed network technologies and provide services independent from access technologies.
- Interconnection and roaming capabilities.

By exposing IMS capabilities to third-party application developers, applications can use the value of an operator community implemented set of functionality within their applications. To enable this, easy-to-use, light-weight, secure and controllable APIs are utilised.

4.9 Further Carrier Aggregation

4.9.1 Technology Overview

Section 3.3 discusses the business opportunity deriving from being able to offer an enhanced mobile broadband service. A short-term solution for operators to increase the data rate that can be offered to consumers is to widen the transmission bandwidth. However, first a technical solution enabling devices to take advantage of this larger bandwidth has to be defined. For this reason, 3GPP introduced in their Release 10 specifications the concept of Carrier Aggregation (CA). Besides making it possible to provide wider bandwidth, CA also allows more flexible load balancing.

4.9.2 Carrier Aggregation Benefits and Requirements

4.9.2.1 Component Carrier

CA works by the aggregation of a number of component carriers (CC) into a wider bandwidth, which is the sum of each CC's bandwidth. Each CC is backwards compatible with R8/R9 LTE carriers. An LTE-Advanced terminal is able to access the entire aggregation of the multiple CCs, thereby realising an overall greater bandwidth and higher data rates. The CCs may be contiguous or non-contiguous, belong to one or different bands. The use of non-contiguous CCs enables spectrum aggregation, i.e. the simultaneous usage of non-contiguous frequency spectrum fragments for a communication to/from a single terminal. In practice, the implementation of spectrum aggregation is constrained and specifications are limited to a number of aggregation scenarios. However, the number of aggregation scenarios are continuously extended in 3GPP. When contiguous CCs exist within the same operating frequency band (as defined for LTE), these are so called intra-band contiguous CCs. Such an arrangement may not always be possible, due to operator frequency allocation scenarios. For non-contiguous allocation, it could either be intra-band, i.e. the CCs belong to the same operating frequency band, but have a gap or gaps in between, or it could be inter-band, in which case the component carriers belong to different operating frequency bands.

As for Release 10, the deployment requirements for carrier aggregation are synchronisation between the aggregated cells and ideal backhaul, which in practice means that the aggregated carrier components are co-located or transmitted from a Radio Remote Head (RRH).

In Release 10 and up to Release 12 Carrier Aggregation, the CC carrier can have a bandwidth of 1.4, 3, 5, 10, 15 or 20 MHz and up to a maximum of five component carriers can be aggregated, hence the maximum aggregated bandwidth is 100 MHz 3GPP specifies radio frequency (RF) requirements in a band-dependent manner for each combination of carriers explicitly, so the actual number of component carriers depends on the specific combinations that are specified.

4.9.2.2 Application of CA

CA may be applied to both Time-Division Duplex (TDD) and Frequency-Division Duplex (FDD) based LTE. In FDD, the number of aggregated carriers can be different in the downlink (DL) and uplink (UL) directions. However, the number of UL component carriers is always equal to or lower than the number of DL component carriers. The individual component carriers can also be of different bandwidths.

Although the (FDD or TDD) Carrier Aggregation feature was specified in Release 10, RF requirements for specific bands have been added progressively along with the different 3GPP RAN4 specifications releases following the evolution of market demand. RF requirements for CA were initially specified in Release 10 for only a few 2CC proof of concept combinations of LTE. Release 10 and 11 focused on 2CC Carrier Aggregation limited to the downlink for the inter-band case. Release 12 introduced the first RF requirements for inter-band 2CC UL CA and 3CC DL CA. The following definitions are applicable to CA combinations:

- Aggregated Transmission Bandwidth Configuration (ATBC): Total number of aggregated physical resource blocks (PRB).
- CA bandwidth class: Indicates a combination of maximum ATBC and maximum number of CCs. In 3GPP Release 10 and Release 11, three such classes were defined:
  - Class A: ATBC ≤ 100 PRBs, maximum number of CCs = 1.
  - Class B: ATBC ≤ 100 PRBs, maximum number of CCs = 2.
  - Class C: 100 PRBs < ATBC ≤ 200 PRBs, maximum number of CCs = 2.
- CA configuration: Indicates a combination of LTE operating band(s) and CA bandwidth class(es), e.g. CA_1C indicates intra-band contiguous CA on LTE operating band 1 and CA bandwidth class C.

3GPP Release 10 defined RF requirements for three CA configurations with many further CA configurations being added in Releases 11 and 12, provided that these spectrum RF requirements are considered release independent. They are indeed applicable to previous releases as long as the CA related feature is supported by the release. For a complete list, see 3GPP TS 36.101 [14].

Each 3GPP Release from Release 11 onwards has provided enhancements to CA:

- Release 11 added multiple timing advance groups, which enabled CCs on cells served by antennas that are geographically separated from the primary CC.
- Release 12 added the capability to aggregate FDD and TDD spectrum.
- Release 13 will include support for up to 32 component carriers, and is looking at the usage of unlicensed spectrum (Licensed Assisted Access) in addition to licensed spectrum for the provision of CA to increase capacity.

4.10 Higher Order MIMO

4.10.1 Technology Overview

In addition to increasing the bandwidth of the mobile transmission, another technique that allows operators to increase throughput and therefore to satisfy the ever-growing demand for mobile broadband and more efficient usage of the limited spectrum is the use of Multiple Input Multiple Output (MIMO). Introduced in 3GPP Release 7 as a means of increasing peak data rates via multi stream transmission for HSPA, MIMO refers to the technique of sending and receiving more than one data stream on a single radio channel at the same time via the use of multiple antennas at both the transmitter and/or the receiver. By exploiting the property of multi-path propagation of de-correlating the data streams even if they use the same bandwidth, MIMO transmissions achieve a better carrier-to-interference ratio at the receiving end.

4.10.2 MIMO Benefits and Requirements

MIMO uses four main performance improvements:

- **Array gain:** This is the power gain that is attained via the use of multiple transmit and/or receiving antennas with respect to a single-input-single-output case.
- **Diversity gain:** The same data is coded and transmitted through different antennas, which effectively increases the power in the channel, increasing cell edge performance at the receiver. The receiver receives N signals and is able to select the strongest signal at any one time.
- **Spatial multiplexing:** The same frequency and modulation used at all transmitters. The sub-streams are independent and can be individually coded. However, total power is not greater than a single transmitter would be as each of the N transmitters transmits with 1/N of the power. Spatial multiplexing is used when the signal-to-interference ratio (SIR) is relatively high.
- **Reduced interference:** The data to be transmitted is intelligently reordered in time (via MIMO space time coding) so that bits lost in one stream are duplicated elsewhere in another stream. This enables the transmitted data to be less susceptible to interference.

MIMO can be used to transmit to multiple users simultaneously (multi-user MIMO or MU-MIMO) or else to a single user (single-user MIMO or SU-MIMO).

Figure 23: 5x20MHz carrier aggregation (3GPP Release 10)

MIMO was initially introduced using 2×2 MIMO (i.e. 2 transmit and 2 receive antennas) with throughputs of up to 250 Mbps. Subsequent enhancements use more sophisticated antenna configurations to provide greater throughput, mobility and efficiency.

When LTE was introduced in 3GPP Release 8, MIMO was included as an integral part of the technology due to its advantages (enabling high data rates to be achieved along with much improved spectral efficiency) far outweighing its disadvantages (added complexity to the system in terms of processing and the increased number of antennas required). 3GPP Release 8 did support enhancements to MIMO such as up to 4x transmission in the downlink for SU-MIMO and significantly improved MU-MIMO performance. However, in 3GPP releases 8 and 9, multiple transmit antennas on the UE were not supported because, in the interests of power reduction, only a single RF power amplifier was assumed to be available.

3GPP Release 10 introduced a number of further MIMO enhancements related to work in LTE-Advanced. These included up to 8× transmission in the downlink for MU-MIMO and up to 4× transmission in the uplink for SU-MIMO, as well as greater peak data rates and improved channel capacity. It is worth noting here that a 4×4 MIMO UE chipset has already been launched and it is expected that a 4×4 MIMO UE will be ready to market in near future. It is also worth mentioning that MU-MIMO in 3GPP Release 10 was designed to be “transparent” – i.e. a UE is not aware whether it is being co-scheduled with other UEs in the same time-frequency resource and thus sees no difference between SU-MIMO and MU-MIMO. This transparency is also utilised at the network side, which uses dynamic SU/MU-MIMO switching to improve the system performance depending on the traffic and channel conditions. This dynamic switching enables the network to choose whether a given UE is best served by SU-MIMO or MU-MIMO.

Beyond LTE-Advanced, there are studies ongoing to initially look at deployment of ever more sophisticated designs and antenna layouts. An example of this is so called Large MIMO or Massive MIMO with tens or even hundreds of antennas used for better performance. The advantages of Massive MIMO are:

- Increased data rate due to greater number of paths being used to realise a greater level of data to be transferred in a given time.
- Increased link signal-to-noise ratio: This is a basic advantage of all MIMO systems but Massive MIMO takes this to a greater level.
- Channel Hardening: The increase in the number of antennas results in the system becoming less sensitive to vagaries in the channel matrix and requiring less signal processing at the receiver.

A key issue as the number of antennas increases is that of antenna placement. As stated previously, for MIMO to operate with satisfaction, the correlation between antennas must be small. This becomes more difficult as the number of antennas increases. There are a number of approaches that can help accommodate a large number of antennas:

- Usage of higher frequencies enables antennas to be more closely grouped in a given physical space (i.e. due to the shortened wavelengths).
- Use of volumetric rather than linear spacing (i.e. 3 dimensions in place of 2 to provide spacing), although this technique may be hampered in some cases (e.g. a modern smartphone is often very thin).
- Use of spatial modulation, which is a technique for reducing transmission complexity in a massive MIMO system.

Additional techniques have also been introduced to improve MIMO performance. One such example is beam forming, which uses linear arrays to enable the antenna to focus on a particular area. This reduces interference and increases capacity as the receiver will have a beam formed in their particular direction. Beam forming is enabled by controlling the phase and amplitude of a waveform at each of the different transmitters to create a resulting pattern of constructive and destructive interference, which creates a beam toward the target UE.

4.11 Heterogeneous Networks

4.11.1 Technology Overview

Further methods to address the growth of data traffic over the mobile networks is to deploy small radio cells where traffic demand is high or to offload the low-value data traffic to Wi-Fi networks. The natural progression of the reduction of cell sizes, offload of data from the central backhaul network and other technologies including the better use of the radio spectrum is generally referred to as Heterogeneous Networks. When considering that the new 5G Radio Access Technology may be designed to operate at very high frequencies and therefore deployed as small cells, heterogeneous networks become a particularly significant enabler not only in the short term, but also in the long term.

4.11.1.1 Operation of Different Cell Sizes

To meet the intensified demand for data capacity and speed levels, mobile operators can expand an existing macro network and densify it by adding more sectors per eNB or deploying more macro eNBs. Reducing the site-to-site distance in the macro network can only be pursued to a certain extent because finding new macro sites becomes increasingly difficult and can be expensive, especially in city centres. An often better alternative is to introduce small cells through the addition of low-power base stations (eNBs, Home eNodeBs (HeNBs) or Relay Nodes) or a Remote Radio Head (RRH) to existing macro eNBs. Site acquisition is easier and cheaper with this equipment, which is also correspondingly smaller.

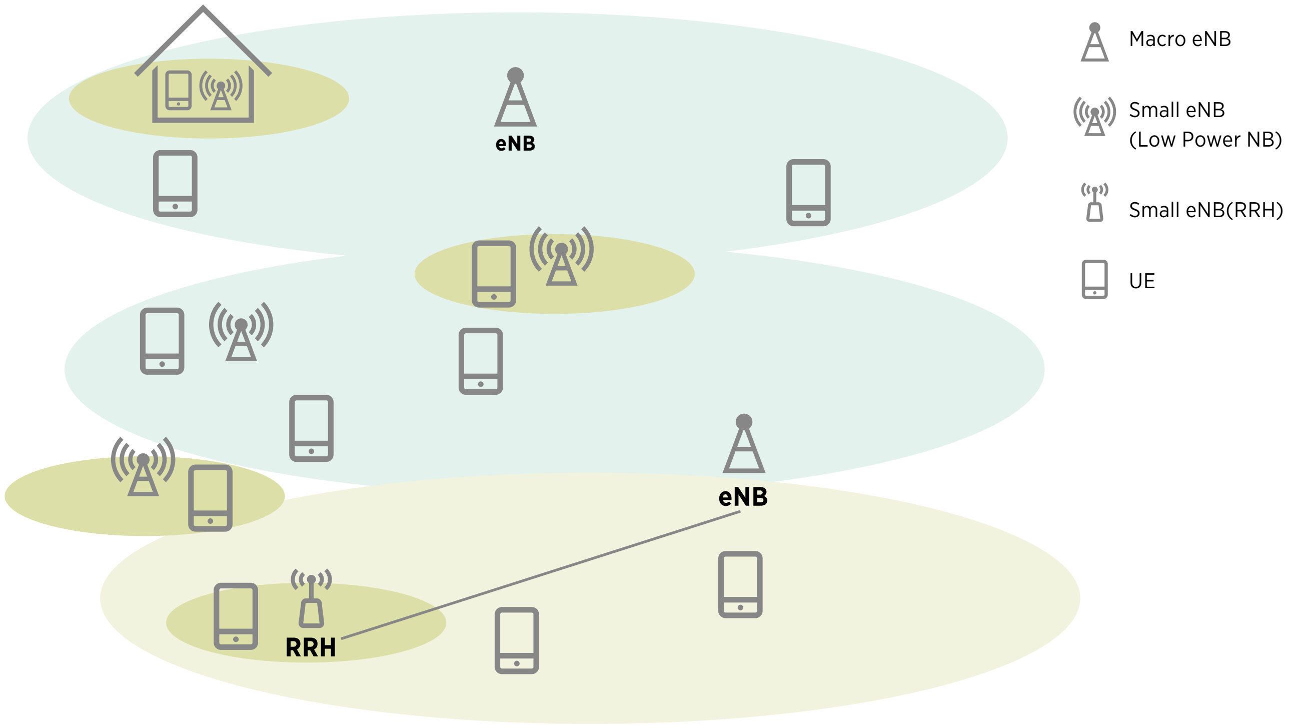
As LTE networks generally use a frequency reuse ratio of 1 to maximise utilisation of the licensed bandwidth, interference between cells causes cell edge performance degradation, which cannot be neglected.

To solve this problem, several solutions have been invented and standardised in 3GPP:

- Further Enhanced Inter-cell Interference Coordination (feICIC) specified in Release 11;
- Cell Range Expansion (CRE);
- cross-carrier scheduling with carrier aggregation; and
- Dual Connectivity.

Source: 3GPP

Figure 24: LTE Heterogeneous Networks



4.11.1.2 Multiple Radio Access Technologies

A different approach to reducing the strain on LTE networks is to offload traffic to other radio access technologies. Offloading traffic to Wi-Fi looks particularly interesting. Wi-Fi can provide significant data capacity, as its carrier bandwidth is very wide (160MHz BW in the latest Wi-Fi specifications), its unlicensed spectrum is very large (580MHz total in 5GHz), it is low power and equipment is relatively cheap thus resulting in easy deployment of numerous cells.

However, Wi-Fi offload presents some challenges when it is desirable to deliver a seamless experience for the user. While 3GPP has specified mechanisms to allow service continuity between 3G and LTE with relative ease due to the fact that both technologies were under its control, the service continuity between Wi-Fi and LTE has been harder to achieve. However, Wi-Fi and LTE interworking has been improved and service continuity can be supported by connecting Wi-Fi to EPC.

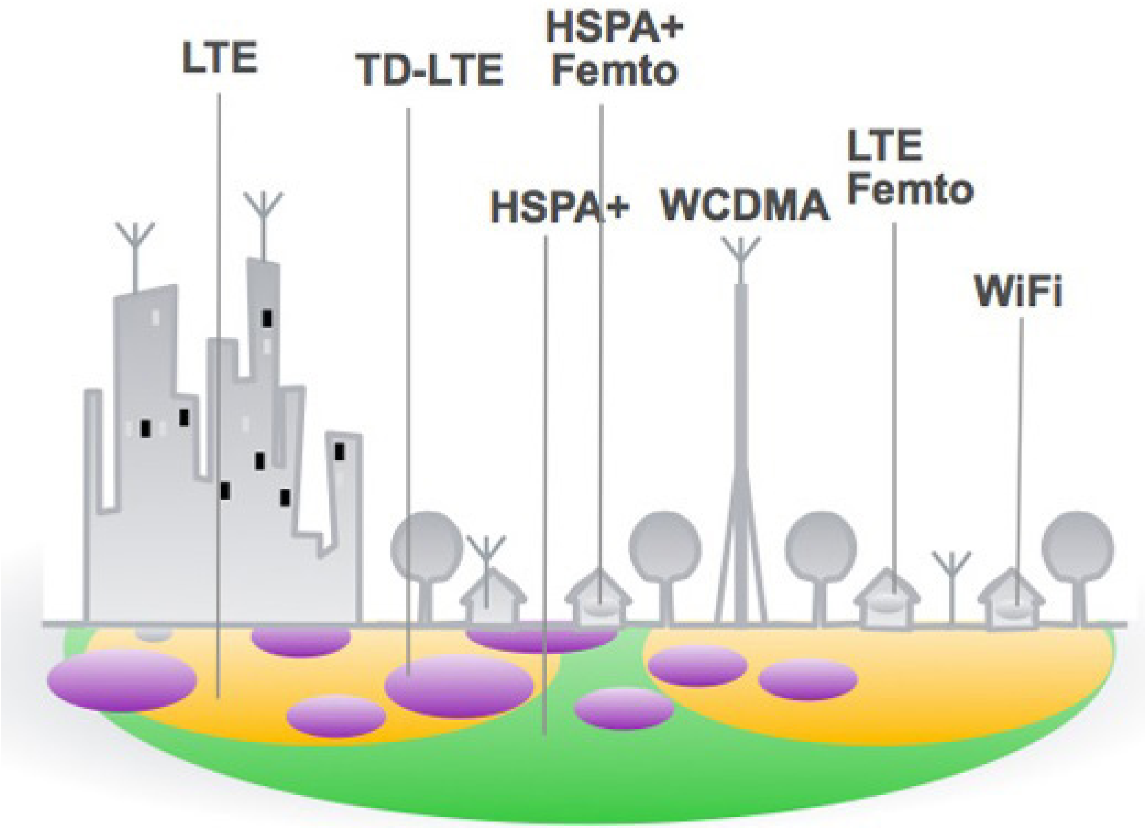
Finally, as any technology can be deployed in the unlicensed spectrum, work is ongoing in 3GPP to define interworking between LTE base stations operating in the licensed spectrum and LTE base stations deployed in the unlicensed spectrum. This activity is known as Licensed Assisted Access LTE (LAA-LTE) and when completed will further enhance the capabilities of heterogeneous networks.

4.11.2 Heterogeneous Networks Benefits and Requirements

As discussed in section 4.11.1, the state of the art for heterogeneous networks is dual connectivity and traffic offload (either to Wi-Fi or using Licensed Assisted Access to LTE). These functionalities are further described in this section.

Source: http://wirelesswire.jp/special/201112/01/article/6.html

Figure 25: Multiple RAT Heterogeneous Networks



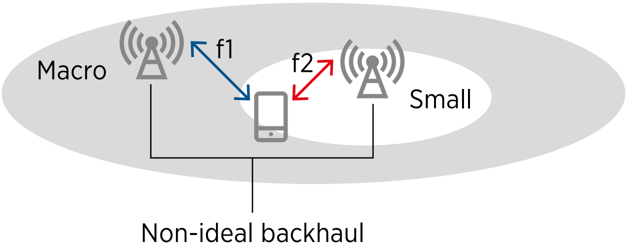
4.11.2.1 Dual Connectivity

The overall concept of dual connectivity (DC) is depicted in Figure 28 where:

- Macro cell node (grey): f1 is typically low frequency e.g. 700M – 2.x GHz, carrying Control Plane mainly.
- Small cell node (purple): f2 is typically high frequency e.g. 3.x GHz or higher, carrying User Plane data mainly (f1=f2 is also possible).
- The macro cell node and the small cell node are connected with non-ideal backhaul.

Source: 3GPP

Figure 26: Overall concept of DC



DC fulfils the following goals:

- Increase UE throughput especially for cell edge UEs.

  DC can significantly increase UE throughput especially for cell edge UEs by transmitting/receiving multiple streams and dynamically adapting to the best radio conditions of multiple cells. Small cells provide additional capacity for UEs having multiple radio connections.
- Mobility robustness enhancement.

  In heterogeneous network deployment, moving UEs can frequently lead to handover failure, inefficient offload and service interruption due to short Time-of-Stay. DC can greatly reduce the handover failure rate by maintaining the macro cell connection as the coverage layer.
- Reducing signalling overhead towards the core network (CN) due to frequent handover.

  This can be achieved by not issuing handover operations as long as the UE is within macro coverage.

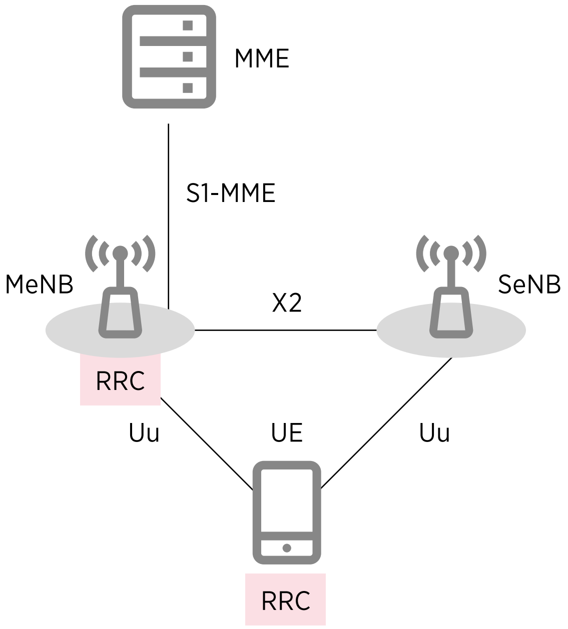
DC is usually configured for low to medium mobility speed cases, and supports indoor and outdoor, ideal and non-ideal backhaul scenarios; however, the focus is on the non-ideal case.

The DC architecture in Release 12 consists of a Master eNB (MeNB) and one Secondary eNBs (SeNB).

The control plane architecture, shown in Figure 29, consists of only one S1-MME connection per UE and it is terminated at the MeNB. The UE Radio Resource Control (RRC) connection is terminated at the MeNB to decrease Radio Resource Management (RRM) and signalling complexity. RRC configurations related to the SeNB are transmitted to the MeNB in the form of an RRC container. Then, the MeNB takes care of all RRC for the UE, constructs the eventual RRC configuration message and transmits it to the UE. The UE always keeps up the RRC connection with the MeNB so long as it exists within the macro cell coverage.

Source: 3GPP

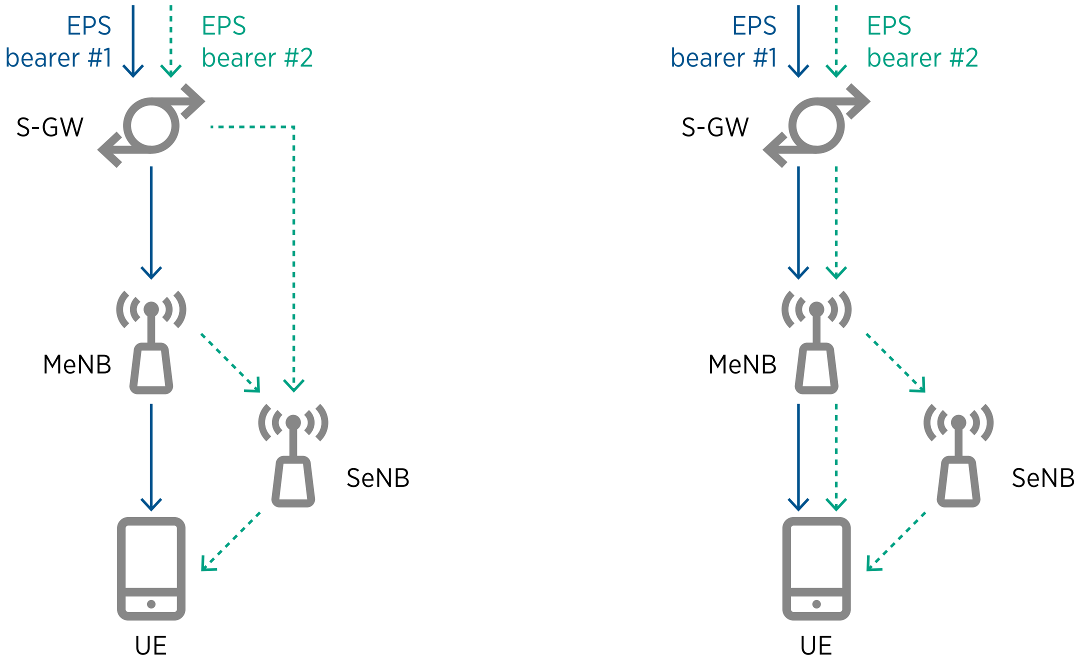
Figure 27: Control plane architecture



With regards to the user plane architecture, two possibilities are available as user traffic may split at the Serving Gateway (SGW) or at the MeNB. It can be debated which user plane architecture is preferable as they both come with pros and cons with regards to impact on the backhaul and on complexity.

Source: 3GPP

Figure 28: User plane architecture



a. SGW routing architecture

b. MeNB routing architecture

It is worth noting that DC may just as well remain applicable where the small cell node utilises the radio access technology defined for use in 5G.

4.11.2.2 LAA-LTE

Existing and new licensed spectrum for exclusive use by International Mobile Telecommunications (IMT) technologies will remain fundamentally critical to providing seamless coverage, achieving the highest spectral efficiency, and ensuring the highest reliability of cellular networks.

In light of the scarcity of bands in frequencies below 6GHz that can be reserved exclusively for IMT technologies, operators are looking at the use of unlicensed spectrum. The latter can never match the qualities of the licensed regime, however, when it is assisted by licensed spectrum, it represents a complementary tool to augment operators' service offering and solutions. In particular, the 3GPP industry is specifying different options to allow that, namely: LAA-LTE and LTE Wireless Local Area Network (LTE-WLAN) aggregation.

Regarding LAA-LTE, the focus is on the use of the 5GHz band where 3GPP started specification work in June 2015 to exploit this opportunity.

LAA-LTE targets the carrier aggregation operation in which one or more low power cells operate in an unlicensed spectrum with assistance from a primary cell that operates in a licensed spectrum. LAA deployment scenarios encompass scenarios with and without macro cell coverage, both outdoor and indoor small cell deployments, and both co-location and non-co-location (with ideal backhaul) between licensed and unlicensed carriers. Figure 31 shows four LAA-LTE deployment scenarios, where the number of licensed carriers and the number of unlicensed carriers can be one or more. Although the backhaul between small cells can be ideal or non-ideal, the unlicensed small cell only operates in the context of carrier aggregation through ideal backhaul with a licensed cell. In scenarios where carrier aggregation is operated within the small cell with carriers in both the licensed and unlicensed bands, the backhaul between macro cell and small cell can be ideal or non-ideal (Scenarios 3 and 4).

Source: 3GPP TR36.889

Table 2: LAA-LTE deployment scenarios (3GPP TR36.889)

Table 2 provides further information about the scenarios depicted in Figure 29:

Term	Description
Scenario 1	Carrier aggregation between a licensed macro cell (F1) and an unlicensed small cell (F3).
Scenario 2	Carrier aggregation between a licensed small cell (F2) and an unlicensed small cell (F3) without macro coverage.
Scenario 3	A licensed macro cell and small cell (F1), with carrier aggregation between the licensed small cell (F1) and an unlicensed small cell (F3).
Scenario 4	A licensed macro cell (F1), licensed small cell (F2) and unlicensed small cell (F3), and carrier aggregation between F2 and F3.

The 3GPP Release 13 specifications on LAA-LTE aim to achieve the following design targets:

- A single global solution framework allowing compliance with any regional regulatory requirements.
- Mandatory, effective and fair coexistence with Wi-Fi in the 5GHz band.
- Mandatory, effective and fair coexistence among LAA networks deployed by different operators in the 5GHz band.
- A minimum set of mandatory features/parameters that enables effective and fair coexistence between Wi-Fi and LAA, and effective and fair coexistence between different LAA systems, should be identified.
- Specify support for LAA small cells operating with only downlink transmissions. 3GPP will decide whether the 5GHz unlicensed band(s) would be specified as a TDD band or an FDD supplemental downlink band or they would specify both possibilities.

Source: 3GPP TR36.889

Figure 29: LAA-LTE deployment scenarios (3GPP TR36.889)

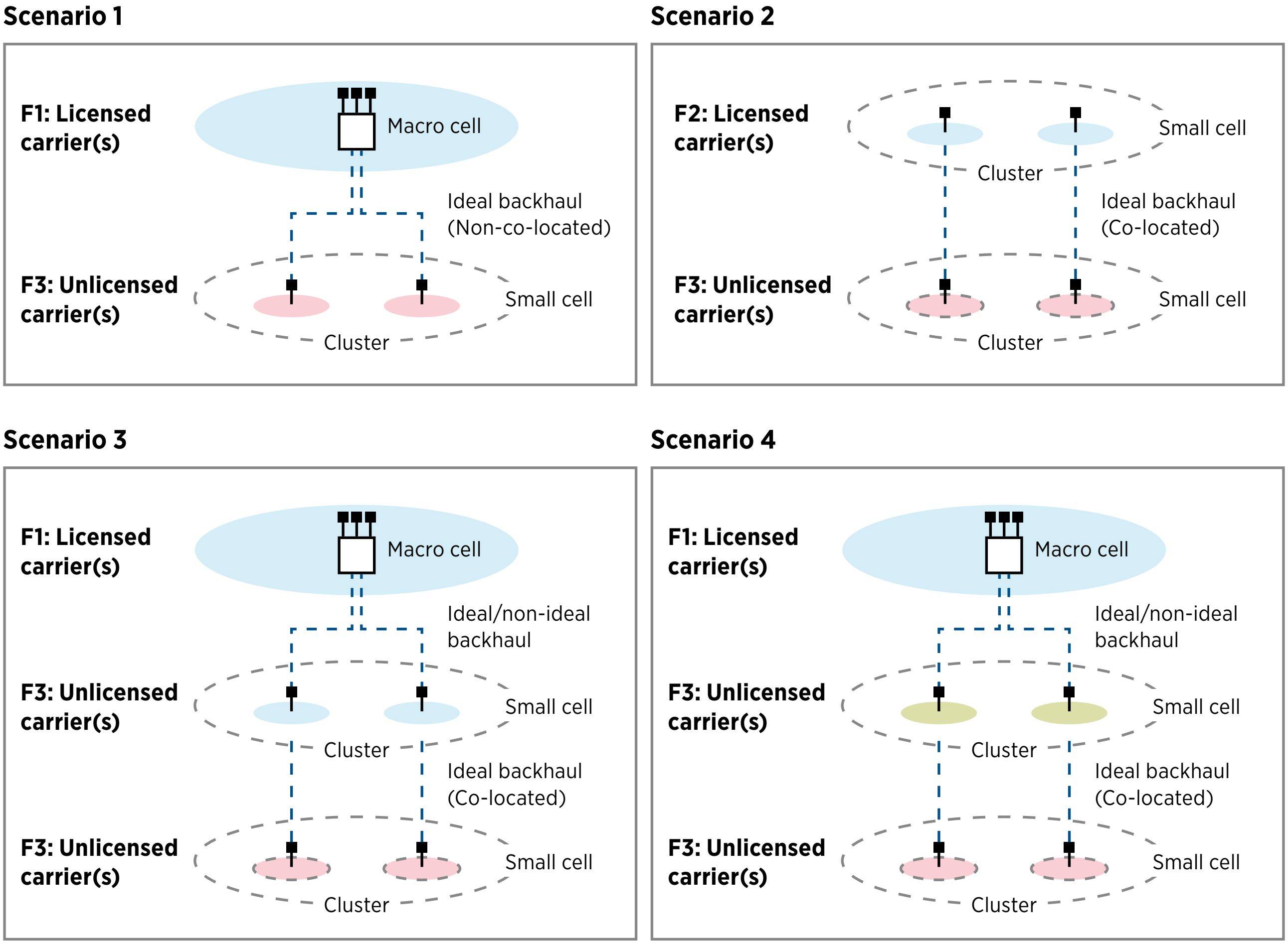
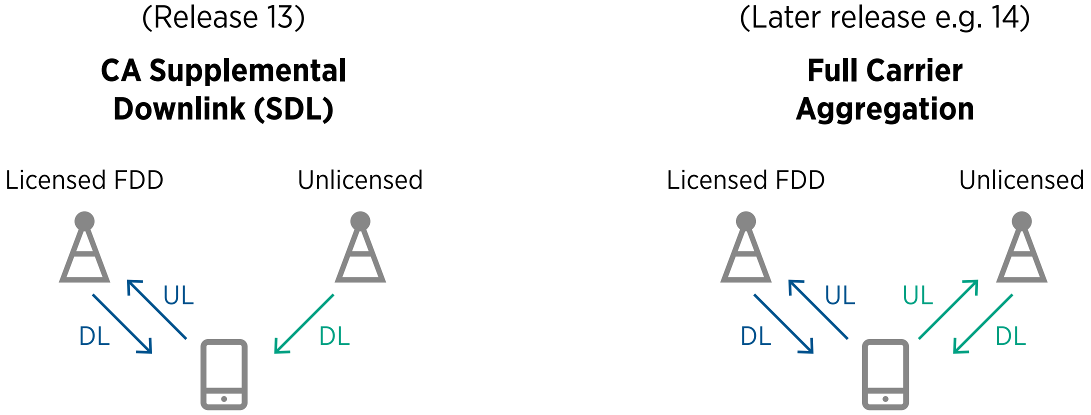


Figure 30 shows the expected LAA-LTE overall configuration of Release 13 and of a future release.

Source: 3GPP

Figure 30: LAA-LTE overall configuration in Release 13 and later

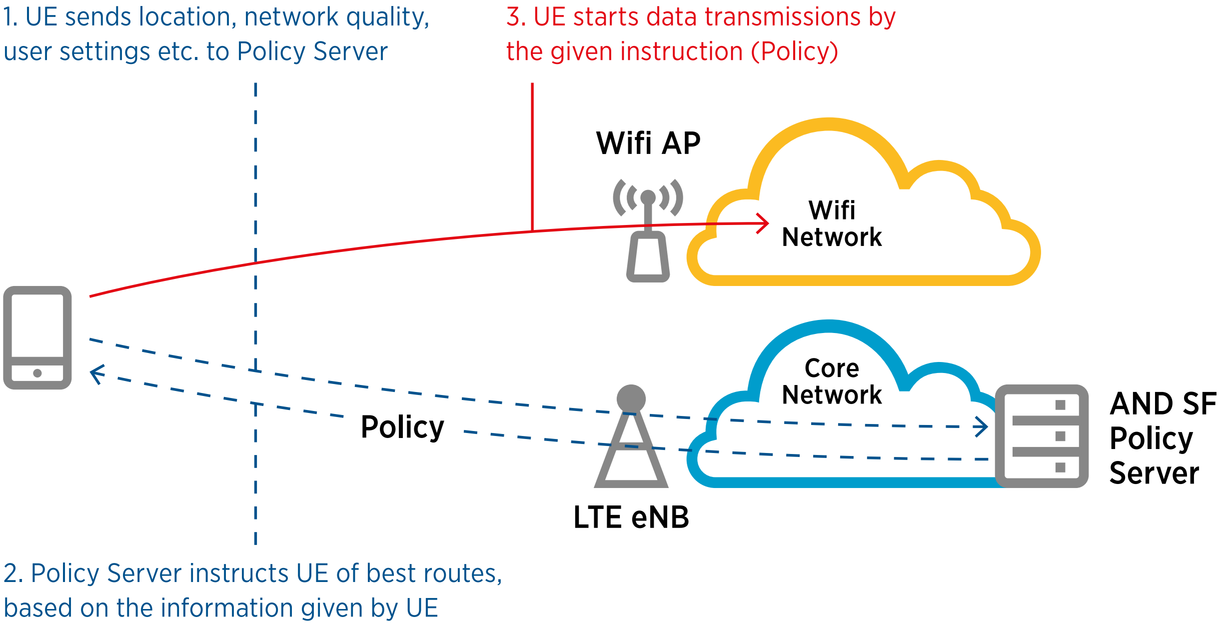


4.11.2.3 Wi-Fi and LTE Interworking

Before Release 12, Wi-Fi interworking with 3GPP technologies was already supported at the CN level, but the Wi-Fi network and PLMN selection mechanisms were not satisfying enough to be deployed, and the Traffic Steering mechanisms via the Access Network Discovery and Selection Function (ANDSF) were not sufficient. Given that different operators have different deployment scenarios, two solutions have been specified in Release 12: an enhanced ANDSF policy based solution; and a RAN rules based solution. If the UE is provided with an ANDSF policy, the UE shall use the ANDSF rule, otherwise the UE may utilise the RAN specified rules.

Source: 3GPP

Figure 31: Wi-Fi/LTE ANDSF policy based interworking



Release 12 introduced enhancements to S2a Mobility over GTP Gateway (SaMOG) enabling UEs to:

- Indicate the requested connectivity type, that is, PDN connection to EPC or non-seamless Wi-Fi offload (NSWO).
- Indicate the APN to establish PDN connectivity.
- Request handover of an existing PDN connection.
- Establish multiple PDN connections in parallel over trusted Wi-Fi.
- Establish an NSWO connection in parallel to PDN connection(s) over Wi-Fi.

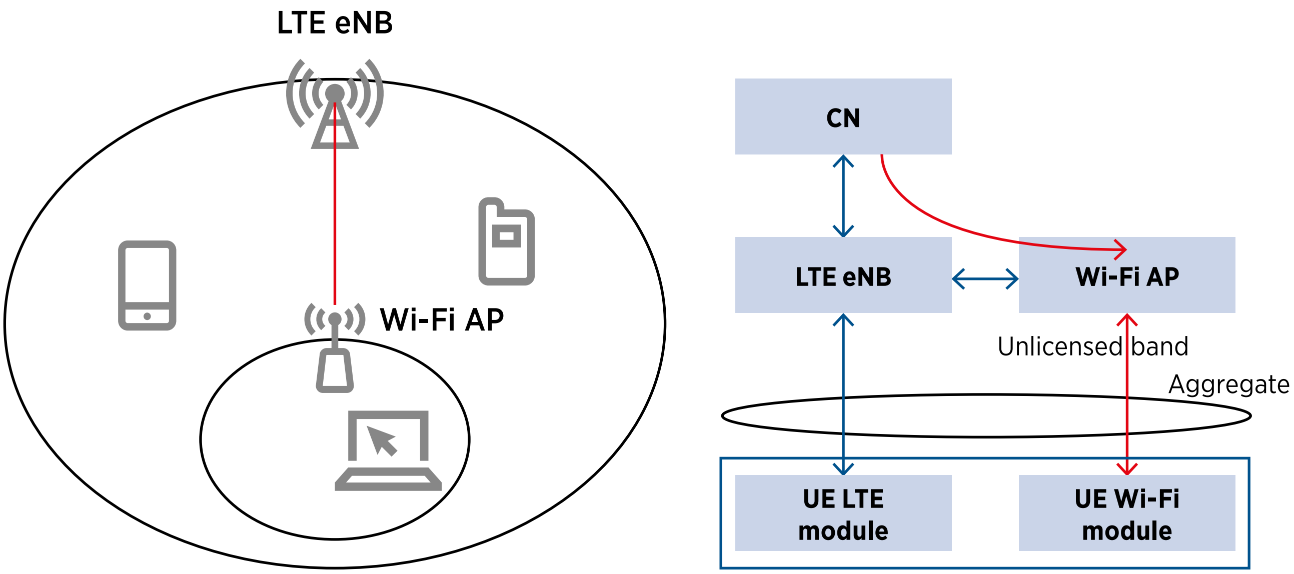
Release 12 also introduced features improving intersystem routing policies and additional clarifications to UE behaviour during inter-RAT mobility addressing bearer loss, QoS degradation and handover ping pong.

In Release 13, further Wi-Fi and LTE interworking is being discussed in 3GPP RAN working groups with an objective to achieve the aggregation of LTE data and Wi-Fi data. Approaches that allow aggregation either at the Packet Data Convergence Protocol (PDCP) layer or above, in the eNodeB are being considered. The two different approaches will have different effects on the network and the terminals. The consequences of aggregation in the eNodeB will have an impact on the capability to collect charging information on the data over the unlicensed spectrum, and the work to define the specifics of eNodeB aggregation is expected to be done in 3GPP in future releases.

As an alternative solution to reduce the dependency on WLAN Access Points, an IP tunnelling solution between the eNB and the UE was proposed. This solution is similar to the evolved Packet Data Gateway (ePDG) based solution with the ePDG functionality moved inside the eNB. However, this solution requires additional study such as IP address allocation for tunnel establishment.

Source: 3GPP

Figure 32: Release 13 interworking – non-co-located scenario



5

Conclusion and Recommendations

This paper discussed how between now and 2020, the year when the 5G system is expected to be commercially available, the technology advancement of the 4G system will continue unabated making it possible to provide existing services more efficiently and start creating a market for emerging applications (e.g. automotive, real-time gaming, personal cloud, sensor networks, remote health monitoring).

Mobile operators can start exploiting business opportunities by evolving their 4G network investments in three areas:

- Enhanced mobile broadband.
- Massive IoT mobilising vertical industries securely.
- Optimised services.

Figure 35 illustrates how already fully specified functionality, ready to be deployed as enhancements to the 4G system, can contribute to improving the services in each of the above mentioned areas.

By providing a detailed description of the 4G technology enablers and the services they enhance, this paper can assist mobile operators in creating their own technical strategic roadmap towards 5G.

Among the enablers described in this document, particular prominence is given to network slicing enabled by advancements in NFV, SDN and orchestration capabilities. Network slicing allows an operator to deploy virtualized instances of networks specialised to serve vertical industries. It is the mechanism through which operators can tailor the same physical hardware to serve not only in the B2C segment, but more importantly to create compelling propositions to vertical industries. Thanks to network slicing and enablers such as Machine Type Communications, operators will no longer be tied by the one-size-fits-all network model deployed today. Compound with high standard security, customisable traffic management and controlled exposure of network capabilities, the evolution of the 4G network will give mobile operators a platform for unlocking commercial opportunities.

Figure 33: Mapping of service enablers to supported service classes

4G evolution enablers		Service classes		
		Enhanced Mobile Broadband	Massive Internet of Things	Optimised Services
System	MEC	✓		✓
	Machine Type Communications		✓	
	Proximity Services		✓	✓
RAN	eMBMS	✓		✓
	Carrier Aggregation	✓		✓
	HetNet (DC, LAA-LTE)	✓	✓	✓
	Massive MIMO	✓		✓
Core Network	Traffic Management	✓		✓
	NFV + SDN	✓	✓	✓

Network
2020

GSMA Head Office
Floor 2
The Walbrook Building
25 Walbrook
London EC4N 8AF
United Kingdom
Tel: +44 (0)20 7356 0600
Fax: +44 (0)20 7356 0601
Email: network2020@gsma.com

Published February 2016